Capacity Underutilization and Demand Driven Business Cycles

Tiancheng Sun*

November 15, 2019

Abstract

I propose a macroeconomic model where firms hold capacity to compete for buyers who are not fully attentive to price differences and thus search for capacity in a somewhat random way. The capacity competition among firms leads to long term capacity underutilization for production. Demand naturally drives business cycles as a result. A shock to consumption demand generates a large movement in investment, an acyclical real wage rate, and a pro-cyclical Solow residual. Estimation results indicate that the single consumption demand shock can already explain most of the business cycle fluctuations.

1 Introduction

Capacity utilization varies substantially over the business cycles and aggregate capacity is never fully utilized. Figure 1 shows the aggregate capacity utilization rate published by the Federal Reserve Board. During a recession, the capacity utilization rate can be as low as 65%. During a boom, it is still likely to be less than 90%. At the micro-level, the Quarterly Survey of Plant Capacity Utilization (QSPC) shows that a large share of plants in the U.S. operate below capacity (Boehm et al., 2017). Similar results are found in the firm-level data for Switzerland (Köberl and Lein, 2011). Like unemployment, the underutilization of capacity suggests that there are always some idle resources in the economy and more so during recessions.

*LSE and Centre for Macroeconomics, Houghton Street, WC2A 2AE, London. Email: t.sun3@lse.ac.uk. Acknowledgments: I thank my supervisor Wouter den Haan for his great support and very detailed comments on earlier drafts. I appreciate helpful comments from participants at LSE Macro work-in-progress seminars and Edinburgh research seminar. I have also benefited from fruitful conversations with John Moore, Ricardo Reis, Nobuhiro Kiyotaki, Mike Elsby, Petr Sedlacek, David Baqaee, Andrew Clausen, Axel Gottfries, Hong Wang, and Lei Liu. All errors are my own.

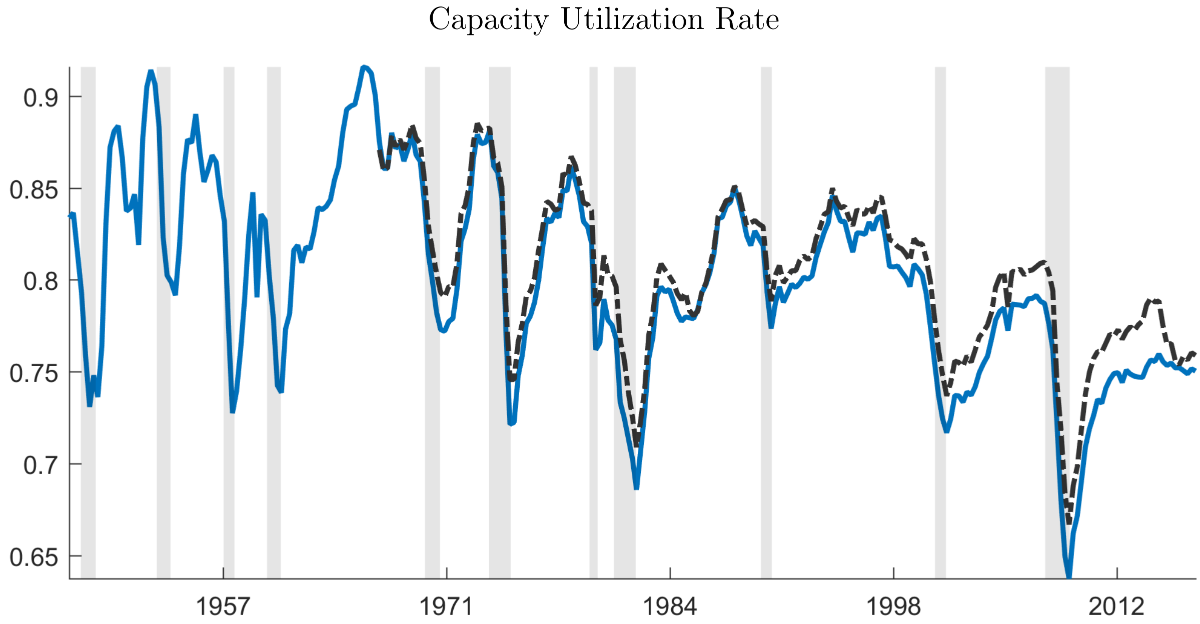

Figure 1: The data is from the Board of Governance of the Federal Reserve System. The solid line is for manufacturing. The dash-dotted line is for total industries. Shaded areas indicate the National Bureau of Economic Research (NBER) dated recessions.

The vast majority of firms in the QSPC cites insufficient demand as the main reason for capacity underutilization (Boehm et al., 2017). Intuitively, when demand goes up, more capacity is utilized; when demand goes down, more capacity is left idle. The observation that demand is important for capacity utilization raises the possibility that demand is important for business cycles.

Recently, demand driven business cycles have received renewed interest, reviving the ideas put forward strongly by Keynes. The standard New Keynesian (NK) literature makes demand important by assuming sticky prices. In most NK models, capacity utilization is not explicitly incorporated. This paper, however, tries to explore the role of demand in a real economy in which prices are perfectly flexible but capacity is generally underutilized.

There is a large literature that incorporates variable capital utilization into macroeconomic models, in which capital is not fully utilized because of a convex capital utilization cost function.[1] However, capital is not used precisely because it is too costly to be utilized. Hence, the standard variable capital utilization model does not really feature idle resources. In fact, I find that *capacity* defined as the output level at which the short run

[1]Greenwood et al. (1988) and Basu and Kimball (1997) assume that the depreciation rate of capital is increasing and convex in terms of the utilization of capital. Kydland and Prescott (1988), Burnside et al. (1993), and Bils and Cho (1994) assume that the overtime premium paid to workers is increasing and convex in terms of the utilization of capital. Smets and Wouters (2007) assume that the amount of intermediate goods used by firms is increasing and convex in terms of the utilization of capital.

average cost curve achieves its minimum, is fully utilized in steady state, even though *capital* has not been fully utilized to work 24 hours a day and 7 days a week.[2] In addition, capital resources are tight as an increase in capital can still reduce the real marginal cost. When demand increases and more capital needs to be utilized, the output to capital ratio goes up and the real marginal cost rises. The upward sloping real marginal cost curve dampens the response of output to demand shocks and causes the real wage rate to be countercyclical under demand shocks. Thus, demand is unlikely to be important in the standard variable capital utilization model.

In this paper, I build a macroeconomic model in which demand turns out to be important for business cycles. In my model, firms hold capacity to compete for buyers. Long term capacity underutilization is an equilibrium result of this capacity competition. When capacity is not fully utilized, capital resources are slack and the real marginal cost curve is flat. As there is no additional cost to utilize idle capacity, output can be highly responsive to demand shocks and the real wage rate is acyclical.

The model has two main assumptions. First, I assume that the production technology is Leontief. Hence, firms can produce with a constant real marginal cost until they are constrained by capacity. Second, I assume that buyers search for capacity to satisfy their demand, and when they process price information, they are subject to an information processing cost as in the rational inattention literature (e.g., Mattsson and Weibull, 2002 and Matějka and McKay, 2015). If the unit cost of processing information is zero, buyers will conduct a directed search and purchase only the cheapest goods; the goods market becomes perfectly competitive. If the unit cost of processing information is infinite, buyers will not process any price information, but conduct an un-directed random search for capacity: firms that have a larger capacity are more likely to be visited. In general, the unit cost of processing information is positive but finite and the behavior of the buyers is somewhere between the directed search and the un-directed random search: firms that charge a lower price and have a larger capacity are more likely to be visited.

Although goods are perfect substitutes, firms can charge positive markups as in the standard Dixit-Stiglitz (DS) market structure because buyers are not fully attentive to

[2] Capital utilization and capacity utilization, though closely related, are *different* concepts. The capital utilization rate measures the number of hours that capital operates relative to the total number of hours within a given period of time. The capacity utilization rate measures the actual level of output within a given period of time relative to capacity. In practice, capacity is considered as the maximum level of output that a firm can produce within a given period of time under a realistic working schedule, taking into account normal downtime (e.g., Corrado and Mattey, 1997). Hence, even if a firm operates at full capacity, the capital of the firm typically works less than 24 hours a day and 7 days a week. In theory, to capture the notion that extraordinary efforts are required to produce beyond capacity, capacity is defined as the output level at which the short run average cost curve achieves its minimum (e.g., Klein, 1960 and Nelson, 1989). Eiteman and Guthrie (1952) conduct a survey and find that the short run average cost curve of a firm is downward sloping until a point near or at capacity. This suggests that the practical notion of capacity is roughly consistent with the theoretical definition of capacity.

prices. What is different from the standard DS market structure is that if one firm expands its capacity relative to that of others, the demand curve faced by the firm shifts outward. In other words, a firm can steal demand from its competitors by expanding its relative capacity. Because firms charge positive markups, it is not only possible but also profitable for firms to steal demand. This puts great pressure on firms to expand capacity. No one wants to be left behind. Long term capacity underutilization is an equilibrium result of this capacity competition.

The inclusion of long term capacity underutilization is important for the model dynamics. As output has not yet reached the capacity limit, capital resources are slack and the real marginal cost curve is flat locally around the steady state according to the assumed Leontief technology. Hence, output can be highly responsive to demand shocks and the real wage rate is acyclical as there is no longer any restriction imposed by capital resource tightness in my model.

This contrasts sharply with the standard real business cycle (RBC) model and the standard variable capital utilization model. In both models, capital resources are tight and the real marginal cost curve is upward sloping. When demand rises, the increase in output is limited because capital as a production factor is scarce, and the real wage rate falls as the marginal productivity of labor falls; similarly, when demand drops, the decrease in output is limited because any capital resource freed up is a valuable production factor that increases the marginal productivity of labor and the real wage rate.

One can reduce the capital resource tightness by reducing the convexity of the utilization cost function in the standard variable capital utilization model. This makes the real wage rate less countercyclical and consumption more responsive. However, the response of investment to consumption demand shocks is inherently small. The reason is the following. When future consumption is expected to increase, firms want to increase their investment precisely because capital resources are tight so that investment can help reduce the future real marginal cost. Thus, if capital resources are less tight, firms will have a weaker desire to invest. In other words, removing the curb on investment also removes the impetus for investment. Hence, it is difficult to get a large relative volatility of investment to consumption and a non-countercyclical real wage rate simultaneously as observed in the U.S. data.

By contrast, in my model, capital resources are slack locally around the steady state and the real wage rate is acyclical. Despite the capital resource slackness, firms still have a strong desire to increase their investment if future consumption is expected to increase. This is because capacity expansion helps a firm to attract demand and the amount of demand that can be attracted by each unit of capacity invested increases if future demand in aggregate increases. Because of this extra motivation for investment, removing the curb on investment does not remove the impetus for investment. Hence, the response of investment to a persistent consumption demand shock in my model is much larger than in standard ones. With capital and investment adjustment costs, the relative volatility of investment to consumption can be consistent with that in the U.S. data.

To assess the importance of demand for business cycles quantitatively in my model, I allow for the possibility that business cycles are driven both by demand shocks and by labor productivity shocks. In particular, I consider three types of demand shocks: a consumption demand shock that changes the marginal rate of substitution between consumption and leisure, an investment demand shock that changes the subjective discount factor, and an exogenous expenditure shock that changes the government expenditure. I use Bayesian estimation techniques to estimate the model.

I find that the model attributes most of the business cycle fluctuations to demand shocks. The labor productivity shock accounts for only 2% of the variation in output and 13% of the variation in hours. Among the three types of demand shocks, the consumption demand shock is the most important. The single consumption demand shock can already explain more than 72% of the variance in consumption, 78% of the variance in investment, and 60% of the variance in hours, and generate the correct business cycles co-movements between consumption, investment, hours, and the Solow residual. These results resonate with two recent empirical papers, which find that business cycles are mainly driven by a single demand shock (e.g., Andrle et al., 2017 and Angeletos et al., 2018).

Summing up, this paper has two major contributions. First, by extending the standard neoclassical framework with buyers who search for capacity and pay an information processing cost, I present a new model that can explain long term capacity underutilization. Second, when viewed through the lens of the model, demand has the potential to be the main driving force of business cycles, and when the model is estimated, that turns out to be the case.

The rest of the paper is organized as follows. The next subsection discusses the related literature. Section 2 establishes a basic capacity underutilization (CU) model. Section 3 studies the properties of the basic CU model. Section 4 compares the basic CU model with a standard variable capital utilization model. Section 5 extends the basic CU model to a full CU model and discusses the quantitative results. Section 6 concludes.

Related Literature

Cooley et al. (1995) and Gilchrist and Williams (2000) assume that plants have different productivity levels. Since low productivity plants are left idle, capacity is not fully utilized. However, in these models, capacity is not fully utilized precisely because it is too costly to be utilized. Hence, if demand increases and plants of low productivity are utilized, the real marginal cost of producing goods will increase, dampening the response of output to demand shocks and causing a countercyclical real wage rate as in the standard RBC model. The same issue is not present in my model, in which capacity is underutilized because of the capacity competition among firms rather than an increased utilization cost.

Fagnart et al. (1997) and Fagnart et al. (1999) assume that firms are subject to large idiosyncratic demand shocks and hold extra capacity as a precaution. In this setup, capacity is on average not fully utilized precisely because there are chances that the capacity constraint is binding. Hence, the idiosyncratic demand shocks must be large enough so that the capacity constraint is frequently binding and a large proportion of firms are running at full capacity each period. Otherwise, firms will not have enough motivation to hold extra capacity as a precaution. In the model, the standard deviation of the year-over-year (YoY) quarterly sales growth rate at the firm-level has to be 60% to generate a capacity utilization rate of 87% on average, and each quarter 47% of firms are running at full capacity. In the real world, the average firm-level volatility of the YoY quarterly sales growth rate is about 15% to 25% (e.g., Comin and Philippon, 2005, Buch et al., 2009, Kelly et al., 2013, and De Veirman and Levin, 2018); and roughly 20% of the firms report running at full capacity each quarter (e.g., Köberl and Lein, 2011 and Boehm et al., 2017). In addition, when aggregate demand increases, more firms have to run at full capacity and the aggregate real marginal cost of producing goods increases, causing a countercyclical real wage rate and dampening the response of output to demand shocks. By contrast, in my model, the real wage rate is acyclical, there is no dampening effect on output, and firms are willing to hold extra capacity to compete for demand even though capacity constraints never bind. Hence, with additional firm-level heterogeneity, my model has a good chance to fit the micro-level data.

Bai et al. (2012) incorporate search and matching into the goods market. They use a competitive search framework as in Moen (1997). Capacity utilization rate is like the market tightness in the standard labor market search and matching literature. Capacity is not fully utilized because it is too costly for buyers to search and purchase goods from a tight market. However, if demand increases and capacity utilization rate increases, the real marginal cost of searching and purchasing goods increases, dampening the response of output to demand shocks. Particularly, the response of investment to a persistent consumption demand shock tends to be too small; thus, they rely on a large positive correlation between consumption demand shocks and investment demand shocks to drive business cycles. In my model, the response of investment to a persistent consumption demand shock is large; thus, a single consumption demand shock already explains most of the business cycles.

Michaillat and Saez (2015) also incorporate search for capacity into the goods market to explain capacity underutilization. They use a random search framework. As a result, firms can use capacity to attract demand as in my model. Their paper, however, focuses on nominal rigidity and studies a preference shock that changes the marginal rate of substitution between consumption and real money balance. My paper focuses on the real economy with flexible prices and studies a preference shock that changes the marginal rate of substitution between consumption and leisure. In their model, they have the freedom to choose the bargaining protocol such that prices are fixed. However, the bargaining process in the goods market is a bit unrealistic as most firms in the real world do have prices quoted in the goods market; and the choice of the bargaining protocol is a bit ad hoc. In my model, firms quote prices flexibly in the goods market to maximize profits.

The literature on industrial organization documents the possibility of established firms holding excess capacity to deter entry (e.g., Spence, 1977, Dixit, 1980, and Bulow et al., 1985). Established firms invest in capacity to protect demand from being stolen by potential entrants. In my model, firms invest in capacity to steal demand from others. Both mechanisms allow capacity to have a positive effect on demand. Because of the complex strategic interactions involved in the entry deterrence literature, this mechanism rarely exists in macroeconomic models. However, the mechanism proposed in this paper can be easily incorporated into the existing macroeconomic framework.

This paper is also related to the operations research and marketing literature. Other things being equal, capacity expansion improves the dependability of a firm and speeds up the delivery process (e.g., Olhager et al., 2001 and Olhager and Johansson, 2012). The improved operational performance increases customer satisfaction (e.g., Zhang et al., 2003, Zhang et al., 2005, Kumar et al., 2011, and Kumar et al., 2013). Finally, customer satisfaction affects the ability of a firm to keep and attract demand (e.g., Rust and Zahorik, 1993, Anderson et al., 1994, and Athanassopoulos, 2000). The assumption that buyers search for *capacity* rather than for firms can be regarded as capturing this phenomenon.

2 A Basic Model of Capacity Underutilization

This section presents a basic capacity underutilization (CU) model. The model has two features. First, the production technology is Leontief. Second, buyers search for capacity to satisfy their demand but they have a limited capability to process price information.

2.1 Technology

There is a unit mass of identical firms indexed by $j \in [0,1]$. All goods produced are perfect substitutes and can be used either as consumption or investment.

At the beginning of each period, each firm has some capital stock inherited from the last period. The law of motion for capital is standard:

$$k_{j,t+1} = k_{j,t}(1-\delta) + i_{j,t}, \tag{2.1}$$

where $i_{j,t}$ is the investment made by firm j in period t, $k_{j,t}$ is the capital stock of firm j at the beginning of period t, and parameter $\delta \in [0,1]$ is the depreciation rate.

The production technology is assumed to be Leontief:

$$y_{j,t} = \min\left\{\frac{l_{j,t}}{\alpha_v}, Ak_{j,t}\right\}, \tag{2.2}$$

where $y_{j,t}$ is the amount of goods produced by firm j in period t, $l_{j,t}$ is the amount of variable labor hired by firm j in period t, parameter $\alpha_v > 0$ is the variable labor required to produce each unit of output, and parameter $A > 0$ is the productivity of capital.

Capacity is defined as the maximum level of output that can be produced by a firm within a given period. The variable labor $l_{j,t}$ is freely adjustable within a given period, while the amount of capital stock $k_{j,t}$ is predetermined. Thus, the capacity of firm j is a linear function of the firm's capital stock:

$$\bar{y}_{j,t} = Ak_{j,t}. \tag{2.3}$$

Because the amount of capital stock is predetermined, capacity may not be fully utilized if demand is not high enough.

I find that it is both theoretically appealing and empirically plausible to start with a simple Leontief production function.

Theoretically, the concept of capacity is naturally clear with a Leontief production function. Thus, papers that explicitly model capacity utilization often assume that firms possess a Leontief type of production function at least in the short run (see, e.g., Fagnart et al., 1997, Fagnart et al., 1999, and Boehm et al., 2017). This is also a standard assumption in the management science and operations research literature.[3]

Empirically, Leontief production functions are not uncommon. For example, drivers and cars are perfect complements in a taxi company; cooks and cookers are perfect

[3] For some papers in the management science and operations research literature, see Florian and Klein (1971), Kalish (1983), and Deng and Yano (2006). For some textbooks on operations management, see Stevenson (2002), Kumar and Suresh (2006), and Gupta and Starr (2014).

complements in a restaurant. At the micro-level, firms are likely to operate with fixed input-output coefficients especially in the short run because these coefficients are much dictated by the technologies embodied in capital, which is carefully designed by modern engineers (Eiteman, 1947).[4] Indeed, numerous empirical evidence based on accounting, engineering, or questionnaire data suggest that marginal cost at the micro-level is typically constant at least up until some point close to capacity (see Walters, 1963, for a literature survey).

2.2 Buyers and the Goods Market Structure

Households and firms do not purchase goods directly. Instead, each household and firm is endowed with a *buyer*. The purchasing process is decomposed into two steps. First, a household or a firm decides how many goods shall be purchased in a given period. Second, the household or the firm sends out her buyer to purchase the goods for her.

If a buyer chooses to purchase goods from firm j, the payoff for the buyer is given by a decreasing function of the price of the goods purchased. In particular, I assume that the payoff function for the buyer is

$$v_t(j) = \ln\left(\frac{1}{P_{j,t}}\right), \tag{2.4}$$

where $P_{j,t}$ is the price charged by firm j.[5]

The buyers are allowed to randomize their choices. The strategy space $\mathcal{F}$ is composed of all the cumulative distribution functions (CDFs) defined on $[0, 1]$.

In principle, a buyer would want to purchase only the cheapest goods that yields the highest payoff. However, as in Mattsson and Weibull (2002), I assume that it is costly for the buyer to implement the desired outcome because she needs to process some price information in order to direct her actions towards the desired.

Without exerting any information processing effort, the buyer can only purchase in a blind and random way. The purchasing behavior in this case is called the *default purchasing behavior*. Let $N_t^* \in \mathcal{F}$ be the default purchasing behavior. I assume that N_t^* is given by an *un-directed random search* for capacity.

The rationale behind the above assumption is as follows. If a buyer pays no attention to prices, the matching between the demand from the buyer and the capacity in the market should be the most *disordered*. Let $f_t(x)$ be the probability density that capacity $x \in [0, \bar{y}_t]$ is matched with demand. The density function f_t describes the matching

[4] This idea is captured by the putty-clay technology introduced by Johansen (1959).

[5] It will become clear later in this section that assuming a log payoff function helps generate a demand curve of constant elasticity.

result. Mathematically, the degree of disorder of the matching result can be measured by the entropy of f_t, which is given by

$$-\int_0^{\bar{y}_t} \ln f_t(x) f_t(x) \, dx.$$

For all distributions with a support limited to the interval $[0, \bar{y}_t]$, the maximum entropy distribution turns out to be the uniform distribution: $f_t(x) = 1/\bar{y}_t$. In other words, the most disordered matching result is that each unit of capacity has an equal probability to be matched.

Hence, the default purchasing behavior of the buyer, which is the most inattentive search process, should be characterized by an un-directed random search for capacity. This implies that the default probability that a buyer shops in firm j is proportional to the capacity of the firm

$$n_t^*(j) = \frac{\bar{y}_{j,t}}{\bar{y}_t}, \tag{2.5}$$

where $n_t^*(j) \equiv dN_t^*(j)/dj$ is the probability density function of N_t^* and $\bar{y}_t \equiv \int_0^1 \bar{y}_{j,t} dj$ is the total amount of capacity in the economy.

Let $N_t \in \mathcal{F}$ be the purchasing strategy eventually obtained by the buyer. To achieve N_t that deviates from the default purchasing behavior N_t^*, the buyer has to process some price information and incur some information processing cost. As in the rational inattention literature, I assume that the information processing cost is proportional to the amount of information gain or disorder reduction measured by the relative entropy of N_t with respect to N_t^*. Relative entropy is also known as Kullback–Leibler (KL) divergence, which is non-negative, convex, and obtains its minimum value zero if $N_t = N_t^*$:

$$D_{KL}(N_t || N_t^*) \equiv \int_0^1 n_t(j) \ln(n_t(j)) - n_t(j) \ln(n_t^*(j)) \, dj,$$

where $n_t(j) \equiv dN_t(j)/dj$ is the probability density function of N_t. Intuitively, the more different N_t is from N_t^*, the more information is needed to be processed.

The buyer's problem can now be characterized as follows:

$$\max_{n_t \geq 0} \int_0^1 v_t(j) n_t(j) \, dj - \Lambda_F \left(\int_0^1 n_t(j) \ln(n_t(j)) - n_t(j) \ln(n_t^*(j)) \, dj \right), \tag{2.6}$$

subject to

$$\int_0^1 n_t(j) \, dj = 1, \tag{2.7}$$

where $\Lambda_F > 0$ is the unit cost of information gain.

Note that $n_t^*(j)$ is exogenous to the buyer's problem because the default purchasing behavior is *not* a rational choice of the buyer.

Mattsson and Weibull (2002) show that the unique solution to the above problem is:

$$\forall j \in [0,1]: \; n_t(j) = \frac{n_t^*(j) e^{\frac{v_t(j)}{\Lambda_F}}}{\int_0^1 n_t^*(j) e^{\frac{v_t(j)}{\Lambda_F}} dj}. \tag{2.8}$$

Equation (2.8) says that the optimal probability of choosing firm j is proportional to the default purchasing probability $n_t^*(j)$ but moderated by the value of purchasing goods from firm j. The higher the value $v_t(j)$ is, the higher the probability that firm j would actually be chosen by the buyer.

We can substitute out $v_t(j)$ using the buyer's payoff function (2.4) and obtain

$$n_t(j) = \frac{n_t^*(j) P_{j,t}^{-\frac{1}{\Lambda_F}}}{\int_0^1 n_t^*(j) P_{j,t}^{-\frac{1}{\Lambda_F}} dj}. \tag{2.9}$$

Since there is a continuum number of buyers, $n_t(j)$ is also the share of the buyers who purchase goods from firm j. Hence, equation (2.9) characterizes how the buyers will be distributed across firms.

If all prices are equal, buyers will be indifferent between choosing any two firms and have no incentive to do any costly information processing. Therefore, the buyers will simply follow an un-directed random search and will be distributed across firms according to the default purchasing probability $n_t^*(j)$.

If prices are different, firms that charge a relatively low price can attract additional buyers. However, because of a positive information processing cost, not all buyers will purchase the goods of the lowest price. Even the goods of the highest price will still be purchased by some buyers, as we can see from equation (2.9) that $n_t(j)$ is always above zero as long as $P_{j,t}$ remains finite. Intuitively, since buyers have only a limited capability of processing price information, they rationally choose to be partially inattentive to prices rather than to purchase always the cheapest goods. Hence, buyers allow themselves to make some "mistakes" with a positive probability in order to save some information processing cost.

Consider a buyer b who wants to purchase $y_{b,t}$ units of goods. The demand that comes from this buyer for firm j is given by

$$y_{b,j,t} = n_t(j)\, y_{b,t}. \tag{2.10}$$

Combined with equation (2.9), the above equation (2.10) becomes

$$y_{b,j,t} = \frac{n_t^*(j) P_{j,t}^{-\frac{1}{\Lambda_F}}}{\int_0^1 n_t^*(j) P_{j,t}^{-\frac{1}{\Lambda_F}} dj} y_{b,t}. \tag{2.11}$$

Let $\varepsilon_D \equiv \frac{1}{\Lambda_F}$ be the elasticity of the demand curve and define an aggregate price index faced by the buyers as

$$P_{I,t} \equiv \left(\int_0^1 n_t^*(j) P_{j,t}^{-\varepsilon_D} dj \right)^{-\frac{1}{\varepsilon_D}}. \tag{2.12}$$

We can rewrite equation (2.11) as a familiar constant elasticity demand function

$$y_{b,j,t} = n_t^*(j) \left(\frac{P_{j,t}}{P_{I,t}} \right)^{-\varepsilon_D} y_{b,t}. \tag{2.13}$$

With a positive information processing cost, buyers act *as if* the homogeneous goods produced by different firms are imperfect substitutes.[6] The lower the information processing cost is, the more competitive the market is. In one extreme where $\Lambda_F \to 0$ and $\varepsilon_D \to +\infty$, buyers are fully attentive to prices, the search and matching process is fully directed, and the goods market is perfectly competitive. In another extreme where $\Lambda_F \to +\infty$ and $\varepsilon_D \to 0$, buyers pay no attention to prices, the search and matching process is fully un-directed and random, and the demand that goes to each firm no longer depends on relative prices.

As is clear from equations (2.13), it is a trivial task to add up the demand from all buyers. The aggregate demand for the goods produced by firm j is

$$y_{j,t} \equiv \int y_{b,j,t} db = y_t n_t^*(j) \left(\frac{P_{j,t}}{P_{I,t}} \right)^{-\varepsilon_D} = y_t \frac{\bar{y}_{j,t}}{\bar{y}_t} \left(\frac{P_{j,t}}{P_{I,t}} \right)^{-\varepsilon_D}, \tag{2.14}$$

where $y_t \equiv \int y_{b,t} db$ is the total amount of goods that the whole economy would like to purchase.

Note that the default purchasing probability $n_t^*(j)$ plays an important role in determining the size of the demand. Although $n_t^*(j)$ is exogenous to the buyer's problem, it is endogenously affected by the relative capacity of the firm: $n_t^*(j) = \bar{y}_{j,t}/\bar{y}_t$. Because of this feature in my model, a firm has *two* ways to increase its demand. First, the firm can lower its relative price as in standard models. Second, the firm can expand its relative capacity because firms with a larger capacity are more likely to be visited by buyers who are not fully attentive to prices and search for capacity in a somewhat random way. For

[6] This helps explain why in reality firms that sell near homogeneous goods, such as oil, steel, and sugar, can still charge a markup to compensate for their fixed costs.

example, Starbucks can expand its sales by opening more brick-and-mortar coffee stores than its competitor Costa. A printing store can expand its sales by installing more printing machines than others. In a standard Dixit-Stiglitz monopolistic competition model, however, $n_t^*(j)$ is assumed to be exogenous to *both* the buyers and the firms, as it is dictated by the preference of a representative household or by the production technology of a final goods sector. Hence, the relative capacity has no effect on the relative demand. Therefore, the demand curve in my model is essentially different from that in the Dixit-Stiglitz setup, though they do share many similarities.

2.3 Households

There is a unit mass of identical households. Consider a representative household who maximizes her expected lifetime utility, which is given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u\left(z_{c,t}, c_t, l_t\right),$$

where $\beta \in [0,1)$ is the subjective discount factor, $z_{c,t}$ is a preference shock, c_t is the amount of consumption goods purchased, and l_t is the labor supply.[7]

The functional form of the instantaneous utility is assumed to be

$$u\left(z_{c,t}, c_t, l_t\right) = \begin{cases} \phi e^{\gamma z_{c,t}} \frac{c_t^{1-\gamma}-1}{1-\gamma} - \bar{\omega} l_t, & \gamma \neq 1 \\ \phi e^{z_{c,t}} \ln\left(c_t\right) - \bar{\omega} l_t, & \gamma = 1 \end{cases} \tag{2.15}$$

where $\phi > 0$ is a scaling parameter, $\gamma^{-1} > 0$ is the elasticity of intertemporal substitution, and $\bar{\omega} > 0$ is the marginal dis-utility of labor. The assumption that the marginal dis-utility of labor is a constant follows the indivisible labor theory proposed by Hansen (1985) and Rogerson (1988).

The preference shock $z_{c,t}$ can be interpreted as a consumption demand shock. An increase in $z_{c,t}$ is an increase in the desire to consume, because it increases the marginal utility of consumption relative to the marginal dis-utility of labor. I assume that $z_{c,t}$ follows an AR(1) process

$$z_{c,t} = \rho_c z_{c,t-1} + e_{c,t}, \tag{2.16}$$

where $\rho_c \in [0,1)$ is a persistence parameter. The disturbance $e_{c,t}$ is independent and identically distributed (i.i.d.) and Normal with mean zero and variance σ_c^2.

Let V_t be the representative household's maximized expected lifetime utility at time

[7] Let $c_t(j)$ be the amount of consumption goods purchased by the household from firm j, we have $c_t(j) = n_t(j)\, c_t$ and $c_t = \int_0^1 c_t(j)\, dj$ according to our assumptions about the buyers.

t. The household's problem at time t can be written as

$$V_t = \max_{\{c_t, l_t\}} \phi e^{\gamma z_{c,t}} \frac{c_t^{1-\gamma} - 1}{1-\gamma} - \bar{\omega} l_t + \beta \mathbb{E}_t (V_{t+1}), \tag{2.17}$$

subject to the resource constraint of the household

$$c_t = w_t l_t + d_t, \tag{2.18}$$

where w_t is the real wage rate and d_t is the amount of real dividends received from firms.

I assume that the labor market is perfectly competitive. Hence, the household takes the wage rate w_t as exogenously given.

The first order conditions (FOCs) of the household's problem are

$$\lambda_t = \phi \left(\frac{c_t}{e^{z_{c,t}}} \right)^{-\gamma}, \tag{2.19}$$

$$\bar{\omega} = \lambda_t w_t, \tag{2.20}$$

where λ_t is the Lagrangian multiplier for the household's budget constraint (2.18) and can be interpreted as the marginal utility of income.

The optimal labor supply condition (2.20) shows that labor supply is infinitely elastic if we hold λ_t constant. In other words, the Frisch elasticity of labor supply is infinite. However, an interior solution is still possible because an increase in the labor supply causes a decrease in the marginal utility of income λ_t (see equations (2.18) and (2.19)).

2.4 Firms

Recall that the capacity of firm j is given by equation (2.3). Thus, we can rewrite the demand function (2.14) for firm j as

$$y_{j,t} = y_t \frac{k_{j,t}}{k_t} \left(\frac{P_{j,t}}{P_{I,t}} \right)^{-\varepsilon_D}, \tag{2.21}$$

where $k_t \equiv \int_0^1 k_{j,t} dj$ is the aggregate capital stock.

The real revenue of firm j is given by $P_{j,t} y_{j,t} / P_t$, where $P_t \equiv \int_0^1 n_t(j) P_{j,t} dj$ is the average price of the goods purchased.

The variable labor required to produce $y_{j,t}$ units of output is $l_{j,t} = \alpha_v y_{j,t}$. Thus, the wages paid by the firm is $w_t \alpha_v y_{j,t}$.

Each period, the firm uses its revenue to purchase investment goods, pay wages, and distribute dividends. Hence, the resource constraint of the firm is given by

$$d_{j,t} + i_{j,t} = \left(\frac{P_{j,t}}{P_t} - w_t\alpha_v\right) y_{j,t}, \tag{2.22}$$

where $d_{j,t}$ is the amount of real dividends paid out by firm j at time t.[8]

Each firm aims to maximize its firm value, which is the present value of the firm's dividend flow. Let $J_{j,t}(k_{j,t})$ be the maximized value of firm j at time t. The firm's problem can be written as

$$J_{j,t}(k_{j,t}) = \max_{\{P_{j,t}, i_{j,t}, d_{j,t}, y_{j,t}, k_{j,t+1}\}} d_{j,t} + \beta\mathbb{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t} J_{j,t+1}(k_{j,t+1})\right), \tag{2.23}$$

subject to the demand curve (2.21), the resource constraint of the firm (2.22), the capacity constraint

$$y_{j,t} \le Ak_{j,t}, \tag{2.24}$$

and the law of motion for capital (2.1).

To prevent firms from charging an infinitely high price, I assume that $\Lambda_F \in (0,1)$. Thus, $\varepsilon_D = 1/\Lambda_F > 1$.

The FOCs of the firm's problem are

$$1 = \beta\mathbb{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\frac{\partial J_{j,t+1}}{\partial k_{j,t+1}}\right), \tag{2.25}$$

$$\frac{P_{j,t}}{P_t} = \frac{\varepsilon_D}{\varepsilon_D - 1}(w_t\alpha_v + \mu_{j,t}), \tag{2.26}$$

where $\mu_{j,t} \ge 0$ is the Lagrangian multiplier of the capacity constraint (2.24).

Equation (2.25) says that the marginal cost of investment should be equal to the expected discounted marginal value of investment in the next period. Equation (2.26) says that the firm should set its price according to a constant markup rule. If the firm operates below its capacity limit, the price charged by the firm will be proportional to the firm's labor cost per unit of output: $w_t\alpha_v$. Once the firm hits its capacity limit, however, the price will be raised up so as to equate the demand $y_{j,t}$ to capacity $\bar{y}_{j,t}$.

Applying the Envelope theorem, we have that the marginal value of capital at the beginning of period t is

$$\frac{\partial J_{j,t}}{\partial k_{j,t}} = \left(\frac{P_{j,t}}{P_t} - w_t\alpha_v - \mu_{j,t}\right) Au_t \left(\frac{P_{j,t}}{P_{I,t}}\right)^{-\varepsilon_D} + A\mu_{j,t} + (1-\delta), \tag{2.27}$$

[8] Let $i_{j,t}(j')$ be the amount of investment goods purchased by firm j from firm j', we have $i_{j,t}(j') = n_t(j')\, i_{j,t}$ and $i_{j,t} = \int_0^1 i_{j,t}(j')\, dj'$ according to our assumptions about the buyers.

where u_t is the aggregate capacity utilization rate defined as

$$u_t \equiv \frac{y_t}{\bar{y}_t} = \frac{y_t}{Ak_t}. \tag{2.28}$$

Equation (2.27) shows that there are two reasons for a firm to invest in capital. First, capital investment relaxes the capacity constraint. Second, capital investment makes the goods produced by the firm more likely to be purchased by buyers.

2.5 Symmetric Equilibrium

In a symmetric equilibrium, if a variable is of the form $x_{j,t}$, we have $x_{i,t} = x_{j,t}$ for all $i, j \in [0, 1]$. After obtaining the FOCs, we can omit the subscripts that index a variable to a particular firm. For example, we have $P_{i,t} = P_{j,t} = P_{I,t} = P_t$ for all $i, j \in [0, 1]$.

The symmetric equilibrium is a stable stochastic process of nine variables (c_t, λ_t, i_t, w_t, μ_t, l_t, u_t, k_{t+1}, and y_t) that satisfies the household's FOCs (2.19)-(2.20), the optimal investment condition

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(\left(1 - w_{t+1}\alpha_v - \mu_{t+1}\right) Au_{t+1} + A\mu_{t+1} + 1 - \delta \right) \right), \tag{2.29}$$

which follows from equations (2.25) and (2.27), the optimal pricing condition

$$1 = \frac{\varepsilon_D}{\varepsilon_D - 1} \left(w_t \alpha_v + \mu_t \right), \tag{2.30}$$

the complementary slackness condition for the capacity constraint

$$\mu_t \left(Ak_t - y_t \right) = 0, \tag{2.31}$$

where $\mu_t \geq 0$ and $Ak_t - y_t \geq 0$, the amount of labor hired according to the Leontief production function

$$l_t = \alpha_v y_t, \tag{2.32}$$

the definition of the capacity utilization rate (2.28), the law of motion for capital

$$k_{t+1} = k_t \left(1 - \delta\right) + i_t, \tag{2.33}$$

and the aggregate resource constraint

$$c_t + i_t = y_t, \tag{2.34}$$

which follows from the resource constraint of the household (2.18) and the resource constraint of the firm (2.22).

3 Properties of the Basic Capacity Underutilization Model

In this section, I show the following properties of the basic CU model. First, the decentralized equilibrium is generally inefficient because of the rational inattention of buyers. Second, if the degree of inattention is large enough, capacity will be underutilized in steady state. Third, locally around the steady state where capacity is underutilized, capital resources are slack. Finally, because of the capital resource slackness, the real wage rate is acyclical and output is highly and much more responsive to demand shocks than in standard RBC models.

3.1 Inefficiency

Given the technology and the preference of my basic CU model, the efficient allocation can be obtained by solving a corresponding social planner's problem.

Consider a social planner who maximizes the lifetime utility of the representative household. Let $\tilde{V}(z_{c,t}, k_t)$ be the maximized lifetime utility when consumption demand is $z_{c,t}$ and capital stock is k_t. The social planner's problem can be written as

$$\tilde{V}(z_{c,t}, k_t) = \max_{\{c_t, i_t, y_t, l_t, k_{t+1}\}} \phi e^{\gamma z_{c,t}} \frac{c_t^{1-\gamma} - 1}{1-\gamma} - \bar{\omega} l_t + \beta \mathbb{E}_t \left(\tilde{V}(z_{c,t+1}, k_{t+1}) \right),$$

subject to the Leontief production function (2.32), the law of motion for capital (2.33), and the aggregate resource constraint (2.34).

The solution to the social planner's problem is a stable stochastic process of eight variables (c_t, λ_t, i_t, μ_t, l_t, u_t, k_{t+1}, and y_t) that satisfies the optimal consumption condition (2.19), the optimal trade-off between the benefit and the cost of production

$$\lambda_t = \bar{\omega}\alpha_v + \lambda_t \mu_t, \tag{3.1}$$

the Euler's equation that gives the optimal investment

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(A\mu_{t+1} + 1 - \delta \right) \right), \tag{3.2}$$

the complementary slackness condition for the capacity constraint (2.31), the required labor followed from the Leontief production function (2.32), the law of motion for capital

(2.33), and the aggregate resource constraint (2.34).

Except for conditions (3.1) and (3.2), the conditions that characterize the efficient allocation are the same as those that characterize the decentralized equilibrium. In the decentralized equilibrium, the optimal trade-off between the benefit and the cost of production is

$$\lambda_t = \frac{\varepsilon_D}{\varepsilon_D - 1} \left(\bar{\omega}\alpha_v + \lambda_t \mu_t \right), \tag{3.3}$$

which follows from equations (2.20) and (2.30), and the Euler's equation that gives the optimal investment is

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(\frac{1}{\varepsilon_D} A u_{t+1} + A\mu_{t+1} + 1 - \delta \right) \right), \tag{3.4}$$

which follows from equations (2.29) and (2.30).

Hence, the equilibrium of the basic CU model is *efficient* if and only if $\varepsilon_D \to \infty$ or $\Lambda_F \to 0$. In this case, the basic CU model is reduced to a standard real business cycle model with a Leontief production technology, or a Leontief-RBC model for short, where buyers are fully attentive to prices and the goods market is perfectly competitive.

In general, as long as buyers are not fully attentive to prices ($\varepsilon_D = 1/\Lambda_F > 1$ or $\Lambda_F \in (0, 1)$), the equilibrium of the basic CU model is *inefficient*.

First, rational inattention of buyers allows firms to enjoy some monopolistic power. Hence, the real wage rate is lowered and the marginal cost of consumption λ_t is inflated by the markup charged by firms. As a result, consumption level tends to be inefficiently low.

Second, when buyers are not fully attentive to prices, capacity also plays an important role in determining the demand allocated to each individual firm (see equation (2.14)). If a firm invests an extra unit of capacity, the firm can expand its own demand. Because of the market power, demand expansion is profitable. However, if all firms enlarge their capacity by the same proportion, the demand allocated to each firm will *not* be affected. Hence, a demand expansion in one firm is achieved at the cost of a demand contraction in others. Because of this externality, capacity holding tends to be inefficiently high.[9]

The first mechanism is usual, which also shows up in a standard Dixit-Stiglitz monopolistic competition model. The second mechanism, however, is new in my basic CU model and helps us understand why firms may hold unused capacity for a long time.

[9]One can see the first mechanism by comparing equation (3.1) with equation (3.3), and the second mechanism by comparing equation (3.2) with equation (3.4).

3.2 Long-term Capacity Underutilization

Let us consider the steady state property of the basic CU model. If a variable is of the form x_t, its value in steady state is denoted by x.

Let $r \equiv 1/\beta - 1$ be the real interest rate in steady state. According to the optimal investment condition of the basic CU model, if $\Lambda_F \in \left(\frac{r+\delta}{A}, 1\right)$, we have

$$u = \frac{r+\delta}{A\Lambda_F} - \frac{\mu}{\Lambda_F} \leq \frac{r+\delta}{A\Lambda_F} < 1,$$

which means that there is long-term capacity underutilization if the degree of inattention to prices is large enough.

If $\Lambda_F \to 0$, however, capacity must be fully utilized in steady state. In this case, the basic CU model is reduced to a Leontief-RBC model and the allocation of the economy becomes efficient. According to the optimal investment condition of the Leontief-RBC model, we have

$$\mu = \frac{r+\delta}{A} > 0,$$

which means that the capacity constraint must be binding in steady state.

The comparison between the basic CU model and the Leontief-RBC model highlights how a small modification on the assumption of the behavior of buyers can cause a substantial change in steady state behaviors.

In the Leontief-RBC model, buyers pay full attention to prices. Thus, relative capacity has no effect on relative demand. The only reason for a firm to invest in capital is to relax its capacity constraint. If the capacity constraint is not binding in steady state, the marginal value of relaxing capacity constraint is zero: $\mu = 0$. The only value for the firm to invest in capital disappears. This cannot be a steady state equilibrium because holding capital is costly. Hence, the firm will reduce its capital holding until its capacity constraint is binding in steady state.

In my basic CU model, however, buyers are not fully attentive to prices, and thus partly search for capacity randomly. Hence, a firm invests in capital not only to relax capacity constraint but also to attract demand. The limited capability of buyers to process price information ($\Lambda_F \in (0, 1)$) also provides firms with a monopolistic power, which allows firms to earn a monopolistic profit for each unit of demand attracted (see equation (2.29)):

$$1 - w\alpha_v - \mu = \frac{1}{\varepsilon_D} = \Lambda_F > 0.$$

The monopolistic profit is a lure for firms to hold extra capacity. Even though capacity is not fully utilized, as long as the monopolistic profit rate due to the rational inattention of buyers is large enough, it is still profitable for firms to expand capacity. This justifies the existence of long-term capacity underutilization in my basic CU model.

Standard models often assume a Cobb-Douglas (CD) production function, with which any output level is feasible within a given period of time. Thus, capacity can no longer be defined as the maximum level of output within a given period of time. To extend the definition to a more general environment, capacity is defined as the output level at which the short run average cost (SRAC) curve is tangent to the long run average cost (LRAC) curve (e.g., Morrison, 1985). Along the LRAC curve, the firm minimizes its average cost with all factor prices evaluated at steady state, adjusting both the variable factors, such as labor, and the short run fixed factors, such as capital. Along the SRAC curve, the firm minimizes the same average cost, adjusting only the variable factors. At the point of tangency, the firm has no incentive to change its capital in the long run.

If the production technology is of constant returns to scale, the LRAC curve is flat. In this case, capacity can also be defined as the output level at which the SRAC curve achieves its minimum. For example, in my basic CU model, the SRAC is

$$\boldsymbol{SRAC}(y,k) = \begin{cases} w\alpha_v + (r+\delta)\frac{k}{y} & y \le Ak \\ \infty & y > Ak \end{cases},$$

and the LRAC is: $\boldsymbol{LRAC}(y) = w\alpha_v + \frac{r+\delta}{A}$. According to the extended definition, capacity $\bar{y}$ is given by Ak, at which the SRAC curve reaches its minimum and is tangent to the flat LRAC curve. This confirms that the new definition is a proper extension.

We are now ready to discuss capacity utilization in a broad class of models. The following proposition illustrates that it is generally difficult to obtain long-term capacity underutilization in standard models without providing an extra motivation for firms to invest in capital to affect demand.

Proposition 1. *Consider a firm who maximizes its firm value subject to a standard law of motion for capital. The demand curve of the firm is represented by $p_t = P(y_t)$, where p_t is the relative price of the firm. $G(y_t, k_t, w_t)$ is the variable cost function of the firm, which is twice differentiable and convex in terms of output and capital. The SRAC with all factor prices evaluated at steady state is given by*

$$\boldsymbol{SRAC}(y_t,k_t) \equiv \frac{G(y_t,k_t,w) + (r+\delta)k_t}{y_t},$$

and the LRAC given by $\boldsymbol{LRAC}(y_t) \equiv \min_{\tilde{k}>0} \boldsymbol{SRAC}\left(y_t, \tilde{k}\right)$ exists.

Suppose that capital expansion always lowers the real marginal cost:

$$G_{yk}(y_t,k_t,w_t) < 0.$$

We have that for each $k_t > 0$, there is a unique output level $\bar{y}_t$ at which the SRAC curve is tangent to the LRAC curve. $\bar{y}_t$ is defined as the capacity of the firm at time t.

In addition, if the demand curve of the firm is NOT affected by the firm's capital, capacity is fully utilized in steady state.

Proof. The FOC of the SRAC minimization problem is $G_k(y_t, k_t, w) + (r + \delta) = 0$. Since the LRAC curve exists and $G_k(y_t, k_t, w)$ is a strictly decreasing function in terms of output: $G_{ky}(y_t, k_t, w) = G_{yk}(y_t, k_t, w) < 0$, there is a unique output level $\bar{y}_t$ that solves the above FOC for each $k_t > 0$.

The firm's problem can be written as

$$\max_{\{y_t, d_t, k_{t+1}\}_{t \geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} d_t,$$

subject to the resource constraint of the firm: $d_t + k_{t+1} = P(y_t)\, y_t - G(y_t, k_t, w_t) + (1 - \delta)\, k_t$.

Since the demand curve $P(\cdot)$ is unaffected by the firm's capital, we have that the optimal investment condition of the firm is

$$1 = \mathbb{E}_t \left(\beta \frac{\lambda_{t+1}}{\lambda_t} \left(-G_k(y_{t+1}, k_{t+1}, w_{t+1}) + 1 - \delta \right) \right).$$

In steady state, the above condition is reduced to $G_k(y, k, w) + (r + \delta) = 0$, which is the FOC of the SRAC minimization problem. In addition, the SRAC is convex in terms of k_t: $G_{kk}(y_t, k_t, w)/y_t \geq 0$. Hence, k minimizes the SRAC at y.

According to the definition of the LRAC, the SRAC curve ($\boldsymbol{SRAC}(\cdot, k)$) is tangent to the LRAC curve ($\boldsymbol{LRAC}(\cdot)$) at the steady state output level y. Let $\bar{y}$ be the capacity that corresponds to the steady state capital level k. The point of tangency that defines capacity for each capital level is unique. Thus, $y = \bar{y}$. We have that capacity is fully utilized in steady state. □

Intuitively, if capital has *no* effect on demand, the only reason for a firm to invest in capital is to reduce cost.[10] However, if output is below capacity, it is capital reduction rather than investment that reduces cost. In this case, the only reason for the firm to invest disappears, but it is costly for the firm to hold capital. Hence, the firm will reduce its capital until its capacity is fully utilized in steady state.[11]

[10]Relaxing capacity constraint can be regarded as a special case of reducing cost. If an output level is infeasible, the production cost at this level is infinite. Thus, relaxing capacity constraint reduces the production cost from infinity to a certain finite number.

[11]Based on the idea of Proposition 1, Hall (1986) finds that the marginal variable cost saved by capital $(-G_k(y, k, w))$ is smaller than the rental cost of capital $(r + \delta)$ in various U.S. industries, and concludes that there is long term capacity underutilization in the U.S..

In my basic CU model, however, long term capacity underutilization is rather natural, as firms hold capacity to compete for demand.

In the rest of the paper, I assume that parameter values are such that the capacity constraint (2.24) never binds in the basic CU model. With this assumption, I get rid of the occasionally binding capacity constraint to focus on the *local* properties of my basic CU model around the steady state with a long-term capacity underutilization.[12]

3.3 Capital Resource Slackness

To explain how the existence of unused capacity affects the model dynamics, I introduce the concept of capital resource tightness (or slackness).

If a marginal decrease in the capital stock leads to an increase in the real marginal cost (RMC) for a given level of output, I say that capital resources are *tight*. If a marginal decrease in the capital stock has no effect on the RMC for a given level of output, I say that capital resources are *slack*.

Capital resource tightness is determined by the negative elasticity of the RMC with respect to capital, which measures the scarcity of capital as a production factor. If the production technology is of constant returns to scale, capital resource tightness also equals to the elasticity of the RMC with respect to output, i.e., the steepness of the RMC curve. Let $\xi_{j,t}$ be the capital resource tightness faced by firm j at time t. We have

$$\xi_{j,t} \equiv -\frac{\partial \ln G_y\left(y_{j,t}, k_{j,t}, w_t\right)}{\partial \ln k_{j,t}} = \frac{\partial \ln G_y\left(y_{j,t}, k_{j,t}, w_t\right)}{\partial \ln y_{j,t}}, \tag{3.5}$$

where $G(\cdot)$ is the variable cost function and G_y gives the RMC.

The upper left panel of Figure 2 illustrates the relationship between the RMC and capital when output and the real wage rate are both evaluated at steady state; the upper right panel of Figure 2 shows the RMC curve where capital and the real wage rate are evaluated at steady state.

In the basic CU model, capacity is underutilized, the RMC curve is flat, and capital resources are slack locally around the steady state. In the Leontief-RBC model, where capacity is fully utilized, the RMC curve is vertical and capital resources are infinitely tight locally around the steady state. Finally, the dotted lines in Figure 2 are for the standard RBC model with a CD production technology, or the CD-RBC model for short, in which capital resources are neither slack nor infinitely tight, but somewhere in between.

[12]It might be interesting to note that if the capacity constraint never binds, one can regard my model as equivalent to a standard monopolistic competition model where the production function is simply linear in labor, unaffected by any physical capital, and firms accumulate a "marketing capital" that affects demand. This analogy breaks down, of course, if the capacity constraint binds occasionally.

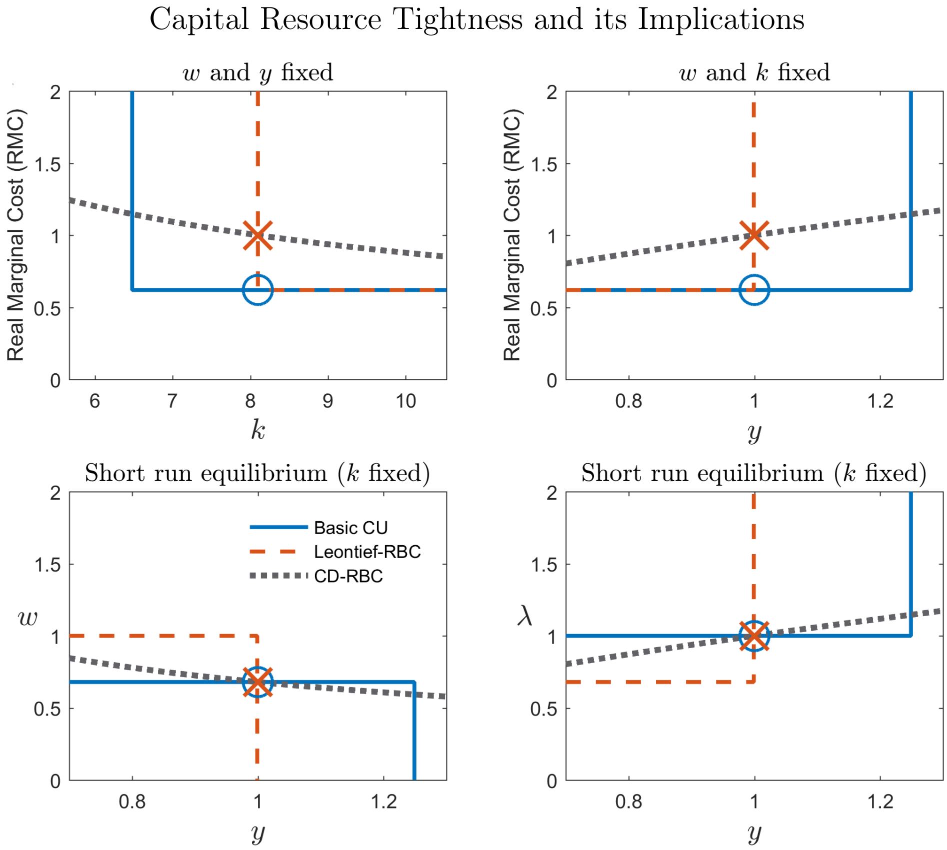

Figure 2: The circle marker is the steady state of the basic CU model and the cross marker is the steady state of the Leontief-RBC model and the CD-RBC model.

3.4 Acyclical Real Wage Rate and Large Responses to Demand

Capital resource slackness implies an acyclical real wage rate and allows output to be highly responsive to demand shocks, while capital resource tightness implies a countercyclical real wage rate and limits the response of output to demand shocks.

In the Leontief-RBC model, capital resources are infinitely tight. When demand increases, output cannot increase, but the marginal value of relaxing the capacity constraint μ_t increases, creating a pressure for the RMC to increase. According to the firm's optimal pricing condition (2.30), this pressure is transmitted to a *decrease* in the real wage rate. The lower left panel of Figure 2 shows this relationship between the real wage rate and output in a short run equilibrium, where capital is fixed and evaluated at steady state. The countercyclical real wage rate will then cause a pro-cyclical marginal utility of income λ_t according to the household's optimal labor supply condition (2.20). The lower right panel of Figure 2 shows this short run equilibrium relationship. As is clear from the household's consumption condition (2.19) and the firm's optimal investment condition (2.29), the pro-cyclical marginal utility of income λ_t reduces the responses of consumption and investment to demand shocks.

By contrast, in the basic CU model, capital resources are slack. If there is a higher demand, more output can be produced without causing any pressure for the RMC to increase. The real wage rate w_t and the marginal utility of income λ_t are both locally constant under the demand shocks (see the solid lines of the lower two panels of Figure 2). Because of the constant marginal utility of income, we do not have any dampening effect in the basic CU model. Hence, consumption and investment can both be highly responsive to demand shocks.

To see this point clearly, I log-linearize the household's optimal consumption condition (2.19) and the firm's optimal investment condition (2.29) of the basic CU model around the steady state where the capacity constraint is not binding. Let $\hat{x}_t \equiv \ln x_t - \ln x$ be the log-deviation of x_t from its steady state x. The log-linearized consumption condition is

$$0 = z_{c,t} - \hat{c}_t, \tag{3.6}$$

and the log-linearized investment condition is

$$0 = \mathbb{E}_t\left(\hat{u}_{t+1}\right) = \mathbb{E}_t\left(\hat{y}_{t+1} - \hat{k}_{t+1}\right), \tag{3.7}$$

where $\hat{k}_{t+1}$ follows a log-linearized capital law of motion: $\hat{k}_{t+1} = (1-\delta)\,\hat{k}_t + \delta\hat{i}_t$.

Nothing dampens the responses of consumption and investment in the basic CU model: equation (3.6) shows that consumption responds one-to-one to the consumption demand shock; equation (3.7) shows that if there is an increase in the expected future output, investment shoots up immediately so that capacity can meet the increased output as soon as possible. Although capital is not scarce as a production factor, firms still have a strong desire to invest because capacity helps attract demand and when future demand is expected to increase the amount of demand that can be attracted by each unit of capacity invested also increases. Since capital resources are slack, all the induced investment can be made immediately in the first period. Therefore, the response of investment to a persistent consumption demand shock in my model is very large.

3.5 Calibration

To illustrate the above results quantitatively, I calibrate the basic CU model at a quarterly frequency.

We have eight parameters to calibrate. The depreciation rate δ is calibrated to match the average ratio of gross private domestic investment to private fixed assets from 1947 to 2016 in the National Income and Product Accounts (NIPA) prepared by the U.S. Bureau

Table 1: Parameters and Calibration Targets – Basic CU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Inter-temporal elasticity of substitution 1
ϕ	0.8300	Marginal utility of income 1
$\bar{\omega}$	0.6798	Output 1
β	0.9747	Capacity utilization rate 0.8
A	0.1544	Investment to output ratio 0.17
α_v	0.9120	Labor underutilization rate 0.088
ε_D	2.6316	Labor share of income 0.62

of Economic Analysis (BEA). The inverse of the inter-temporal elasticity of substitution γ is chosen to be 1, a value that implies a commonly used log utility.

The rest of the six parameters are jointly calibrated to achieve the following targets in steady state.

The marginal utility of income in steady state λ is normalized to 1. This target is mostly associated with the scaling parameter ϕ in the representative household's utility function.

The size of output in steady state y is normalized to 1. This target is mostly associated with the dis-utility of working parameter $\bar{\omega}$, which affects the size of the economy through the supply of labor.

The investment to output ratio in steady state i/y is matched to the average ratio of gross private domestic investment to gross domestic product (GDP) from 1947 to 2016 in NIPA. This target is mostly associated with the productivity of capital A, which affects the capital to output ratio.

The labor underutilization rate in steady state is defined as one minus the ratio of the labor hours actually utilized to the total labor hours that the representative household can potentially supply: $1 - l$, where the total hours that the representative household can supply is normalized to one. I choose the average of U-5 and U-6 from 1994 to 2016 prepared by the Bureau of Labor Statistics (BLS) as a target for the labor underutilization rate.[13] This target is mostly associated with the required labor per unit of output α_v, which determines the demand for labor.

The capacity utilization rate in steady state u is matched to the average of the total industry capacity utilization rate from 1967 to 2016 reported by the Federal Reserve

[13] According to BLS, U-5 is defined as total unemployed, plus discouraged workers, plus all other persons marginally attached to the labor force, as a percent of the civilian labor force plus all persons marginally attached to the labor force. U-6 is defined as total unemployed, plus all persons marginally attached to the labor force, plus total employed part time for economic reasons, as a percent of the civilian labor force plus all persons marginally attached to the labor force.

Board (FRB). This target is mostly associated with the subjective discount factor β, which affects the opportunity cost of holding capacity.

The labor share of income in steady state wl/y is matched to the average labor share of income estimated by BLS from 1946 to 2016. This target is mostly associated with the demand elasticity ε_D, which affects the size of the monopolistic profit and thus the labor share of income.

Table 1 summarizes the calibrated parameter values and their mostly associated targets.[14]

3.6 Impulse Responses and Discussions

Figure 3 plots the impulse response functions (IRFs) for the basic CU model together with those for the Leontief-RBC model and those for the CD-RBC model. The first two rows of Figure 3 show the case where the demand shock is completely transitory ($\rho_c = 0$), while the last two rows of Figure 3 show the case where the demand shock is highly persistent ($\rho_c = 0.99$).

Indeed, although the only difference between the Leontief-RBC model and the basic CU model is the assumption on the behavior of buyers, the local dynamics of the two models are drastically different.

In the Leontief-RBC model, the capacity constraint is binding and capital resources are infinitely tight. Output is restricted by capacity and cannot be changed immediately. Hence, consumption and investment must move in opposite directions. If there is a one-off increase in consumption demand ($\rho_c = 0$), a 1% increase in consumption demand can only lead to a 0.84% increase in consumption, and the increase in consumption must be satisfied by a sharp decrease in investment. If the increase in consumption demand is highly persistent ($\rho_c = 0.99$), it is worthwhile to invest in capital so that households can enjoy a higher consumption level in the long run. However, the induced investment must be satisfied by a temporary decrease in consumption. Besides, the real wage rate correlates negatively with demand shocks and the Solow residual is acyclical.[15] In the U.S. data, consumption and investment co-move, the real wage rate is weakly pro-cyclical or acyclical, and the Solow residual is strongly pro-cyclical. These illustrate the typical difficulties of having demand to be the main driving force of business cycles in an economy where all capacity is fully utilized.

[14]For comparison, I have also calibrated the Leontief-RBC model and the CD-RBC model. In these two models, since capacity is fully utilized in steady state, I do not use the observed average capacity utilization rate as a calibration target. The other calibration targets are the same as in the basic CU model. See Appendix A for details.

[15]Throughout this paper, the Solow residual is calculated by assuming a CD production function for the final products and a CD labor share of 0.62.

Impulse Responses for the Basic CU Model, the Leontief-RBC Model, and the CD-RBC Model

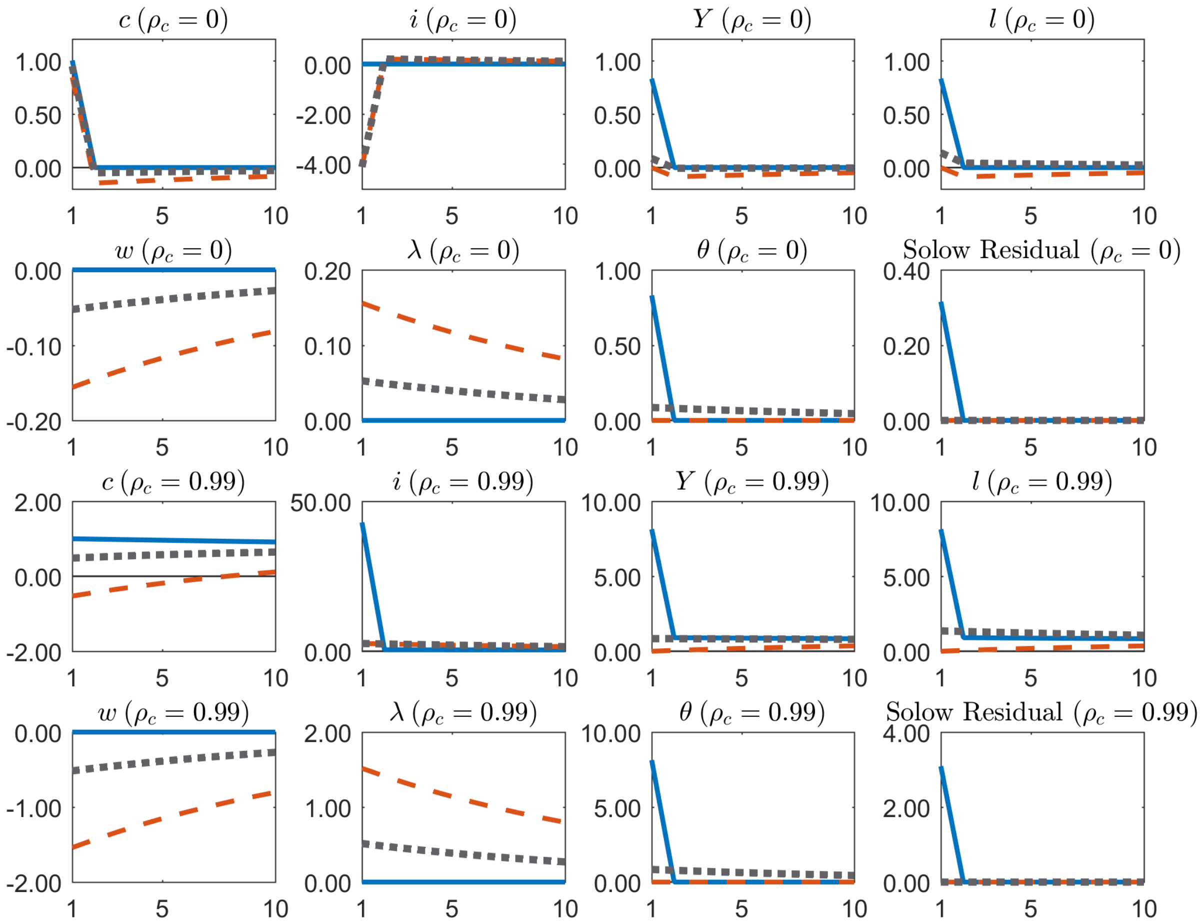

Figure 3: Responses to a 1 percentage point increase in demand disturbance e_c. The solid lines are for the basic CU model. The dashed lines are for the Leontief-RBC model. The dotted lines are for the CD-RBC model. All variables are expressed as log deviations from steady state.

In the basic CU model, however, capacity is underutilized and capital resources are slack. Consumption and investment can now move in the same direction. If demand shocks are completely transitory ($\rho_c = 0$), a 1% increase in consumption demand can lead to a 1% increase in consumption and *no* investment has to be sacrificed.[16] If demand shocks are persistent ($\rho_c = 0.99$), a 1% increase in consumption demand can lead to a 1% increase in consumption and also a huge increase in investment. Firms want to increase

[16]The one-off response of consumption does not mean that consumption smoothing motive is absent. Consumption smoothing motive does not say that a household wants her consumption to be stable but says that she wants her marginal utility of consumption to be stable. Since the demand shock is a preference shock that increases the marginal utility of consumption $\phi e^{\gamma z_{c,t}} c_t^{-\gamma}$ only in the first period, to smooth out the marginal utility of consumption across time, the best thing that the household can do is to consume more in the first period to offset the effect of the increased $z_{c,t}$ and then to consume normally as $z_{c,t}$ goes back to normal. This is exactly what she does in the equilibrium of the basic CU model.

their investment because the amount of demand that can be attracted by each unit of capacity invested increases when future consumption demand is expected increase. A 1% increase in future consumption requires roughly a 1% increase in capacity, which in turn requires roughly a $(1/\delta)\,\% = 48\%$ increase in investment. Since capital resources are slack, investment can be made immediately so that capacity is raised up as soon as possible to compete for the increased future demand.[17] Because of the huge induced investment, a 1% persistent increase in consumption demand is amplified to more than 8% increase in output. In addition, the capital resource slackness also allows the real wage rate to be acyclical. Finally, as a result of a pro-cyclical capacity utilization, the Solow residual is highly pro-cyclical under demand shocks.

In the CD-RBC model, capital resources are tight but not as tight as in the Leontief-RBC. Hence, the response of output to demand shocks can be larger than in the Leontief-RBC model, but is still much smaller than in the basic CU model. Particularly, the response of investment to a persistent consumption demand shock is as limited as in the Leontief-RBC model. In addition, the real wage rate correlates negatively with demand shocks and the Solow residual is acyclical.

To sum up, capital resource slackness, together with the strong motivation of firms to invest in capacity to compete for demand, is the main reason why output can be highly responsive to demand shocks in my basic CU model. Particularly, one can get a positive co-movement between consumption and investment, a large fluctuations in investment, an acyclical real wage rate, and a pro-cyclical Solow residual, even though all fluctuations are driven by demand shocks. These results are difficult to be obtained in standard RBC models. Thus, demand shocks are more likely to drive business cycles in my model than in standard ones.[18]

4 Comparison with a Standard Variable Capital Utilization Model

In this section, I introduce a standard model that features a variable capital utilization, or a standard VU model for short, in which capital is not fully utilized because of a convex

[17]This lack of persistence in the response of investment in the basic CU model follows from the capital resource slackness caused by the existence of unused capacity, but does not depend on the linear relationship between the capacity and capital. Suppose that the production function of the firm is given by $y_{j,t} = \min\left\{l_{j,t}/\alpha_v, Ak_{j,t}^{\alpha}\right\}$, where $\alpha \in (0,1)$. In this example, even though capacity is given by a nonlinear function of capital, $\bar{y}_{j,t} = Ak_{j,t}^{\alpha}$, we will still have the response of investment to demand shocks concentrated in the first period.

[18]A drawback of the basic CU model is that the response of investment lacks persistence. This issue, however, can be easily resolved by an inclusion of capital and (or) investment adjustment costs (see section 5.1, for further discussions).

utilization cost function. Despite having variable *capital* utilization, I show that *capacity* is fully utilized in steady state and capital resources are tight. As a result, the real wage rate is countercyclical and the response of output to demand shocks is dampened.

One can reduce capital resource tightness by reducing the convexity of the utilization cost function. Although this makes the real wage rate less countercyclical and consumption more responsive, I show that the response of investment to a persistent consumption demand shock is inherently small. Hence, it is difficult to achieve a large relative volatility of investment to consumption as observed in the U.S. data without causing a strongly countercyclical real wage rate. Therefore, the role of demand in driving business cycles is likely to be much limited in the standard VU model.

4.1 The Standard VU Model

4.1.1 Setup

The setup of the households is exactly the same as in the basic CU model (see section 2.3).

The goods market is of a standard Dixit-Stiglitz monopolistic competition structure. There exists an aggregate good produced as a composition of individual goods

$$y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon_D - 1}{\varepsilon_D}} dj \right)^{\frac{\varepsilon_D}{\varepsilon_D - 1}}, \tag{4.1}$$

where $y_{j,t}$ is the amount of good j, y_t is the amount of the aggregate good, and $\varepsilon_D > 1$ is the elasticity of substitution between any two individual goods.

Each individual good is produced by a single firm indexed by $j \in [0,1]$. Households and firms demand only the aggregate good. Thus, the demand curve faced by firm j is given by

$$y_{j,t} = y_t \left(\frac{P_{j,t}}{P_t} \right)^{-\varepsilon_D}, \tag{4.2}$$

where P_t is the aggregate price defined as $P_t \equiv \left(\int_0^1 P_{j,t}^{1-\varepsilon_D} dj \right)^{\frac{1}{1-\varepsilon_D}}$.

Each firm produces goods with a CES production function

$$y_{j,t} = F\left(l_{j,t}, \vartheta_{j,t} k_{j,t}\right) = \left(\alpha \left(\frac{l_{j,t}}{\alpha_v} \right)^{\frac{\epsilon_k - 1}{\epsilon_k}} + (1-\alpha)\left(\theta_{j,t} A k_{j,t}\right)^{\frac{\epsilon_k - 1}{\epsilon_k}} \right)^{\frac{\epsilon_k}{\epsilon_k - 1}}, \tag{4.3}$$

where $\theta_{j,t}$ is the capital utilization rate, $\alpha \in (0,1)$ is the CD weight of labor, and $\epsilon_k \in [0, \infty)$ is the elasticity of substitution between capital and labor. It can be shown that when $\epsilon_k \to 0$, we are back to the Leontief production function as in the basic CU model;

and when $\epsilon_k \to 1$, we obtain the usual CD production function:

$$y_{j,t} = \left(\frac{l_{j,t}}{\alpha_v}\right)^{\alpha} (\theta_{j,t} A k_{j,t})^{1-\alpha}. \tag{4.4}$$

Following Smets and Wouters (2007), the cost of utilizing capital is proportional to the size of capital stock, and increasing and convex in terms of the capital utilization rate: $a(\theta_{j,t}) k_{j,t}$, $a'(\theta_{j,t}) > 0$, and $a''(\theta_{j,t}) > 0$.

It is convenient for us to decompose the firm's problem into two steps.

1. Cost Minimization. For a given level of output and capital, the firm minimizes its variable cost by choosing labor input and capital utilization rate

$$G(y_{j,t}, k_{j,t}, w_t) = \min_{l_{j,t}, \theta_{j,t}} w_t l_{j,t} + a(\theta_{j,t}) k_{j,t}, \tag{4.5}$$

subject to its production constraint $F(l_{j,t}, \theta_{j,t} k_{j,t}) \geq y_{j,t}$. It can be shown that the variable cost function G is twice differentiable, homogeneous of degree one, and convex in terms of output and capital. In addition, an increase in capital lowers the real marginal cost (RMC): $G_{yk} < 0$.[19]

2. Firm Value Maximization. Taking the variable cost function G as given, the firm maximizes its firm value. Let $J_{j,t}(k_{j,t})$ be the maximized value of firm j at time t. The firm value maximization problem can be written as:

$$J_{j,t}(k_{j,t}) = \max_{\{P_{j,t}, d_{j,t}, y_{j,t}, k_{j,t+1}\}} d_{j,t} + \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} J_{j,t+1}(k_{j,t+1}) \right) \tag{4.6}$$

subject to the demand curve (4.2) and the resource constraint for dividends

$$d_{j,t} + k_{j,t+1} = \frac{P_{j,t}}{P_t} y_{j,t} - G(y_{j,t}, k_{j,t}, w_t) + (1-\delta) k_{j,t}. \tag{4.7}$$

It is interesting to note that the Leontief-RBC model and the CD-RBC model can be regarded as limit cases of the standard VU model. Particularly, let

$$\bar{a}(\theta_{j,t}) \equiv \begin{cases} 0 & \theta_{j,t} \leq 1 \\ \infty & \theta_{j,t} > 0 \end{cases}.$$

If $a(\cdot) \to \bar{a}(\cdot)$, $\varepsilon_D \to \infty$, and $\epsilon_k \to 0$, the standard VU model converges to the Leontief-RBC model; if $a(\cdot) \to \bar{a}(\cdot)$, $\varepsilon_D \to \infty$, and $\epsilon_k \to 1$, the standard VU model converges to the CD-RBC model.

[19] For a proof of the properties of G, see Appendix C.

4.1.2 Symmetric Equilibrium

The symmetric equilibrium is a stable stochastic process of eight variables (c_t, λ_t, w_t, i_t, θ_t, k_{t+1}, l_t, and y_t) that satisfies the household's FOCs (2.19)-(2.20), the firm's optimal pricing condition

$$1 = \frac{\varepsilon_D}{\varepsilon_D - 1} G_y \left(\frac{y_t}{k_t}, 1, w_t \right), \tag{4.8}$$

the firm's optimal investment condition

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(G_y \left(\frac{y_{t+1}}{k_{t+1}}, 1, w_{t+1} \right) \frac{y_{t+1}}{k_{t+1}} - G \left(\frac{y_{t+1}}{k_{t+1}}, 1, w_{t+1} \right) + 1 - \delta \right) \right), \tag{4.9}$$

the firm's optimal capital utilization condition

$$G_y \left(\frac{y_t}{k_t}, 1, w_t \right) F_K \left(l_t, \vartheta_t k_t \right) = a' \left(\theta_t \right), \tag{4.10}$$

the law of motion for capital (2.33), the production function (4.3), and the aggregate resource constraint

$$c_t + i_t + a \left(\theta_t \right) k_t = y_t, \tag{4.11}$$

where $a(\theta_t) k_t$ is the amount of intermediate goods used as a cost of capital utilization. Let Y_t be the real value added or the amount of final goods produced. We have

$$Y_t = y_t - a \left(\theta_t \right) k_t = c_t + i_t \tag{4.12}$$

in the standard VU model.[20]

4.2 Properties of the Standard VU Model

4.2.1 Inefficiency

The decentralized equilibrium of the standard VU model is efficient if and only if the goods market is perfectly competitive ($\varepsilon_D \to \infty$).

If the goods market is imperfectly competitive ($\varepsilon_D > 0$), firms can set a markup and earn a monopolistic profit. However, unlike firms in the basic CU model, firms in the standard VU model cannot expand capacity to compete for buyers. Without capacity competition, the monopolistic profit alone does not generate long-term capacity underutilization nor capital resource slackness.

[20] In the basic CU model, the Leontief-RBC model, and the CD-RBC model, we do not have any intermediate goods used; thus, the amount of final products is simply given by $Y_t = y_t = c_t + i_t$.

4.2.2 No Long-term Capacity Underutilization

Capacity is defined as the output level at which the short run average cost reaches its minimum. Since the firm's demand curve (4.2) is not affected by the firm's capital stock, Proposition 1 can be applied and there is *no* capacity underutilization in the steady state of the standard VU model.

Intuitively, when capacity has no effect on demand, the only reason for firms to invest in capital is to reduce cost. If capacity is underutilized in steady state, the short run average cost curve is downward sloping and it is capital reduction rather than investment that reduces cost. Hence, firms will decrease their capital stock until their capacity is no larger than sales in steady state. In other words, there is no capacity underutilization in steady state.

4.2.3 Capital Resource Tightness

An increase in capital increases the marginal labor productivity and lowers the real marginal cost (RMC): $G_{yk} < 0$. Thus, by definition, capital resources are tight.

Because the production function (4.3) is of constant returns to scale, capital resource tightness can also be measured by the steepness of the RMC curve. As capacity is fully utilized in steady state, the short run average cost curve is flat at steady state and the RMC curve must be upward sloping locally around the steady state. The magnitude of the capital resource tightness in steady state is given by

$$\xi = (1-\alpha)\,\frac{\xi_\theta^a}{1+\alpha\epsilon_k\xi_\theta^a} > 0,$$

where $\xi_\theta^a \equiv \frac{a''(\theta)\theta}{a'(\theta)} > 0$ is the convexity of the capital utilization cost function in steady state.

The more convex the capital utilization cost function is, the tighter the capital resources. The upper panels of Figure 4 illustrate how the capital resource tightness is affected by the convexity of the capital utilization cost function in a standard VU model with a CD production function ($\epsilon_k = 1$) and a perfectly competitive market ($\varepsilon_D \to \infty$), or a CD-VU model for short. When capital utilization cost function becomes more convex, a decrease in capital causes a larger increase in the RMC and the RMC curve becomes steeper.

4.2.4 Countercyclical Real Wage Rate and Limited Responses to Demand

Capital resource tightness implies a countercyclical real wage rate and limits the response of output to demand.

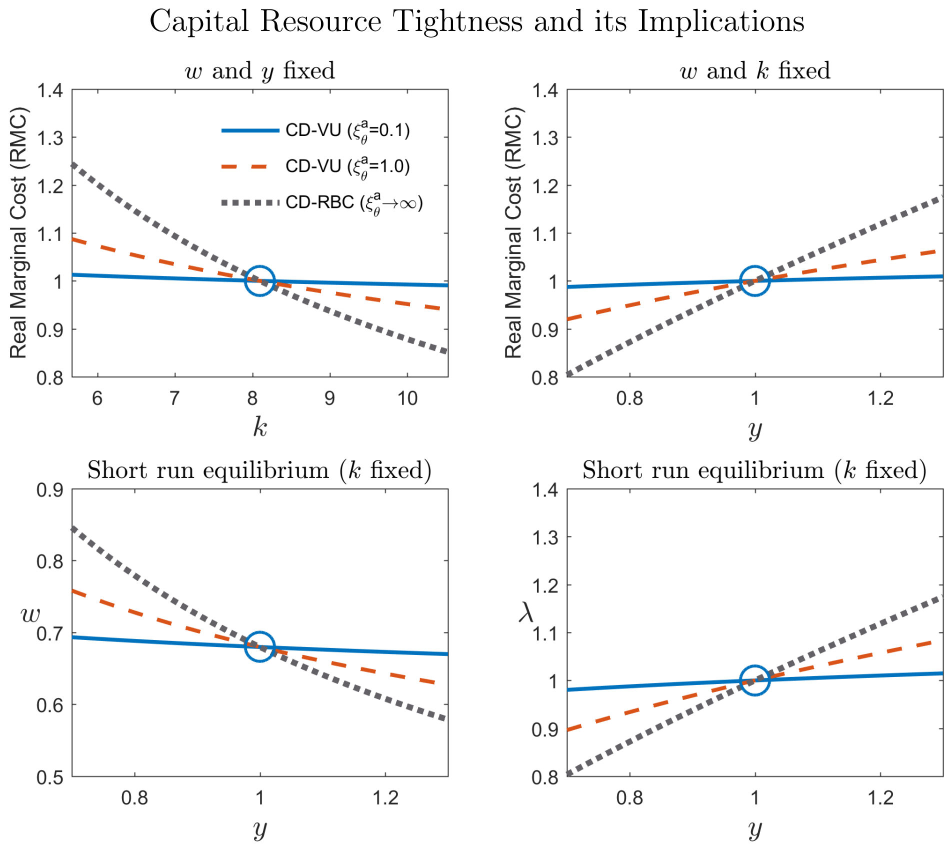

Figure 4: The circle marker indicates the steady state of the CD-VU models.

In the standard VU model, since capital resources are tight, an increase in output generates a pressure for the RMC to increase. This pressure will be transmitted to a decrease in the real wage rate according to the firm's optimal pricing condition (4.8). The lower left panel of Figure 4 shows this relationship between the real wage rate and output in a short run equilibrium. At a first order approximation, we have

$$\hat{w}_t = -\zeta \left(\hat{y}_t - \hat{k}_t \right),$$

where $\zeta \equiv \frac{1-\alpha}{\alpha} \frac{\xi_\theta^a}{1+\epsilon_k \xi_\theta^a} > 0$. The countercyclical real wage rate will then cause a pro-cyclical marginal utility of income λ_t according to the household's optimal labor supply condition (2.20). The lower right panel of Figure 4 shows this short run equilibrium relationship. Finally, the pro-cyclical marginal utility of income reduces the responses of consumption and investment to demand shocks.

To see this point clearly, I log-linearize the household's optimal consumption condition (2.19)

$$\zeta \left(\hat{y}_t - \hat{k}_t \right) = \hat{\lambda}_t = \gamma \left(z_{c,t} - \hat{c}_t \right), \tag{4.13}$$

and the firm's optimal investment condition (4.9)

$$\zeta\left(\hat{y}_t-\hat{k}_t\right)=\hat{\lambda}_t=\mathbb{E}_t\left(\left(1+\beta\frac{wl}{k}\right)\zeta\left(\hat{y}_{+1}-\hat{k}_{t+1}\right)\right). \tag{4.14}$$

Equations (4.13) and (4.14) show that the pro-cyclical marginal utility of income λ_t reduces the responses of consumption and investment.

If the magnitude of the capital resource tightness converges to zero ($\xi \to 0$), we have $\zeta \to 0$ for all $\alpha \in (0,1)$ because

$$\xi\frac{1}{\alpha}\geq\xi\frac{1+\alpha\epsilon_k\xi_\theta^a}{\alpha+\alpha\epsilon_k\xi_\theta^a}=\zeta>0.$$

In this case, the real wage rate is almost acyclical and consumption is almost as responsive as in the basic CU model.

One might think that investment can also be almost as responsive as in the basic CU model if the magnitude of the capital resource tightness is negligible. This conjecture, however, is not true. Suppose that the capital resource tightness is reduced. On one hand, the marginal utility of income λ_t is less pro-cyclical so the dampening force becomes weaker (see the left hand side of equation (4.14)). On the other hand, investment becomes more sensitive to changes in the marginal utility of income λ_t (see the right hand side of equation (4.14)). These two effects cancel out. Therefore, the response of investment in the standard VU model is always much limited.

Intuitively, firms want to increase their investment precisely because capital resources are tight so that investment can reduce the future real marginal cost when future consumption is expected to increase. If capital resources become less tight, firms will have a less desire to invest; thus, a small increase in the required rate of return due to the increased marginal utility of income λ_t leads to a large decrease in investment.[21] In other words, removing the curb on investment also removes the impetus for investment. Hence, the response of investment to a persistent consumption demand shock is inherently small in the standard VU model.

4.3 Calibration

To document the above results quantitatively, I calibrate a CD-VU model, where $\epsilon_k = 1$ and $\varepsilon_D \to \infty$. Since capacity is fully utilized in steady state, I do not use the observed average capacity utilization rate as a calibration target. Also note that capital produc-

[21]Techincally, if capital resources become less tight, the curvature of the variable cost function with respect to capital becomes smaller. Thus, capital investment has a smaller effect on the marginal return of capital, which is the real marginal cost saved by capital investment. Hence, a larger change in investment is needed in order to adjust the marginal return on capital to match the changed required rate of return caused by the change in the marginal utility of income λ_t.

Table 2: Parameters and Calibration Targets – CD-VU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Inter-temporal elasticity of substitution 1
ϕ	0.8300	Marginal utility of resource 1
$\bar{\omega}$	0.6798	Output 1
β	0.9747	Investment to output ratio 0.17
α	0.6200	Labor share of income 0.62
α_v	0.9120	Labor underutilization rate 0.088
A	0.1235	$y/(Ak)$ normalized to 1
θ	1.0000	θ normalized to 1
$a(\theta)$	0.0000	$a(\theta)$ normalized to 0

tivity A is not identified because for each value of $A > 0$, there exists a value of α_v such that the total factor productivity (TFP) $\alpha_v^{-\alpha} A^{1-\alpha}$ is the same. I choose A such that $y/(Ak)$ is normalized to 1. Capital utilization rate in steady state θ is normalized to 1. As in Smets and Wouters (2007), the capital utilization cost in steady state $a(\theta)k$ is normalized to 0. The convexity of capital utilization cost function measured by ξ_θ^a is set to be 0.1, 1, or infinity, for a sensitivity analysis. The other calibration targets are the same as in the basic CU model.

Table 2 lists the calibrated parameter values and their mostly associated calibration targets.

4.4 Impulse Responses and Discussions

Figure 5 shows the IRFs for the CD-VU model with different degrees of capital resource tightness. Generally speaking, despite having a variable capital utilization, the dynamic properties of the CD-VU model are not too different from those of the standard CD-RBC model.

If the convexity of the capital utilization cost function ξ_θ^a converges to infinity ($\xi_\theta^a \to \infty$), the CD-VU model converges to the CD-RBC model, in which case capital resources are quite tight, the real wage rate is strongly countercyclical, and the response of output to demand shocks is much limited.

If ξ_θ^a is smaller, capital resources are less tight, the real wage rate is less countercyclical, and consumption is more responsive. However, regardless of the magnitude of the capital resource tightness, the response of investment to demand shock is always as limited as in the CD-RBC model, which is dwarfed when compared to that in the basic CU model (see Table 3 for a comparison).

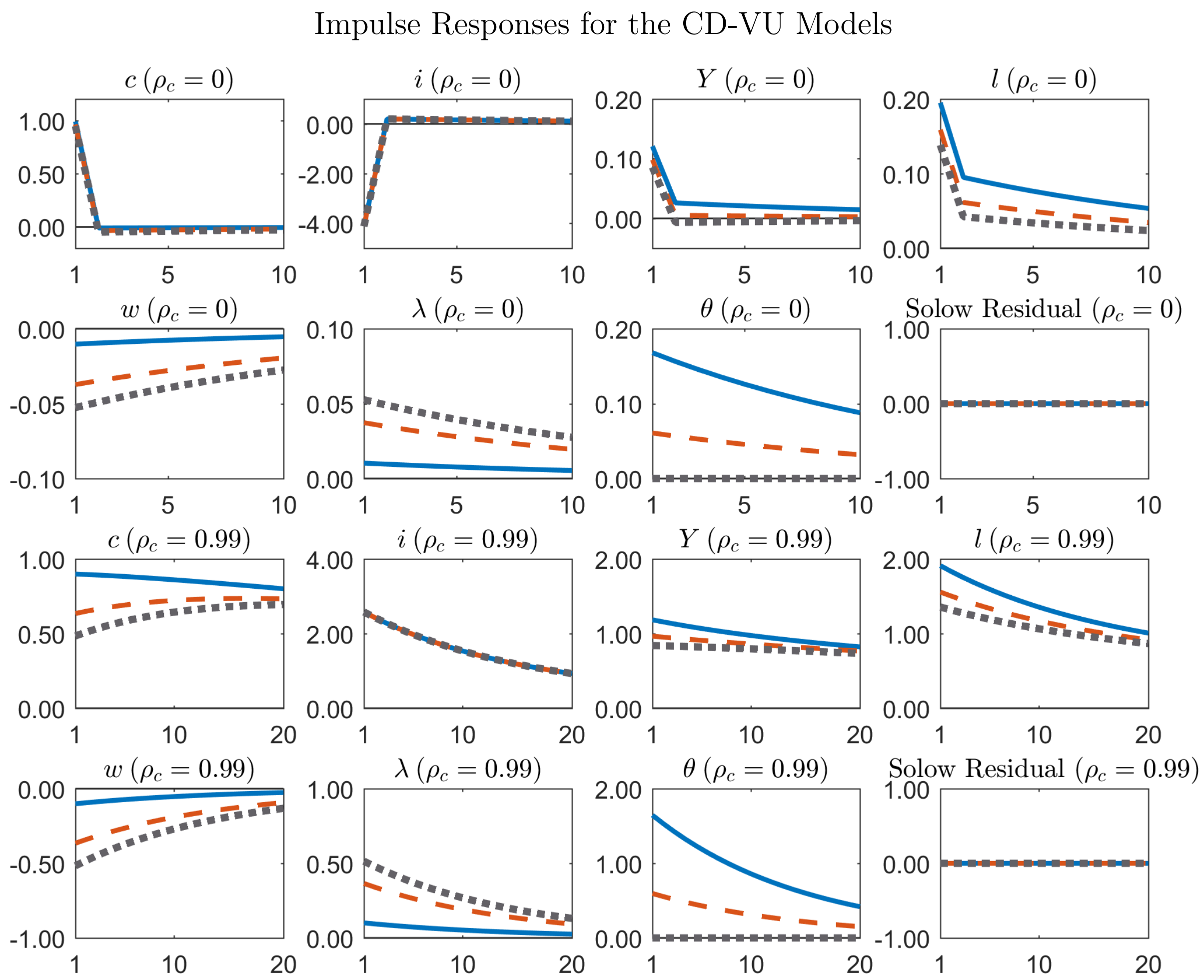

Figure 5: Responses to a 1 percentage point increase in demand disturbance e_c. The solid lines are for the CD-VU model with $\xi_\theta^a = 0.1$. The dashed lines are for the CD-VU model with $\xi_\theta^a = 1$. The dotted lines are for the CD-RBC model or the CD-VU model with $\xi_\theta^a \to \infty$. All variables are expressed as log deviations from steady state.

Since consumption is more responsive but investment is not when capital resources are less tight, the relative volatility of investment to consumption *decreases* when capital resource tightness decreases. As a result, to achieve a large relative volatility of investment to consumption as in the U.S. data, capital resources must be tight enough so that the volatility of consumption relative to that of investment is small. In this case, however, the real wage rate must be highly countercyclical. Hence, the standard VU model fails to generate a large relative volatility of investment to consumption without causing a strongly countercyclical real wage rate.

Table 3 shows that among the CD-VU models with different degrees of convexity of the capital utilization cost function ξ_θ^a, the standard CD-RBC model ($\xi_\theta^a \to \infty$) actually has the highest relative volatility of investment to consumption, which, however, is still a bit smaller than that observed in the U.S. data.[22]

Finally, the Solow residual is always acyclical, despite a pro-cyclical capital utilization rate.[23]

To sum up, in the standard VU model, capacity is fully utilized in steady state and capital resources are tight. Unlike my basic CU model, the standard VU model has a difficult time generating a large movement in investment, an acyclical real wage rate, and a pro-cyclical Solow residual under demand shocks. These issues limit the role of demand shocks in driving business cycles in the standard VU model.

Table 3: Initial Responses and Relative Volatility of Investment to Consumption

	$\rho_c = 0$				$\rho_c = 0.99$			
	c	i	Y	σ_i/σ_c	c	i	Y	σ_i/σ_c
Basic CU	1.00%	0.00%	0.83%	0.00	1.00%	43.0%	8.13%	31.63
Leontief-RBC	0.84%	-4.12%	0.00%	4.57	-0.53%	2.58%	0.00%	4.36
CD-RBC	0.95%	-4.12%	0.09%	4.34	0.48%	2.58%	0.84%	5.07
CD-VU ($\xi_\theta^a = 1$)	0.96%	-4.12%	0.10%	4.31	0.63%	2.58%	0.97%	3.99
CD-VU ($\xi_\theta^a = 0.1$)	0.99%	-4.12%	0.12%	4.24	0.90%	2.58%	1.18%	2.86
The U.S. data				6.49				6.49

Note: Initial responses to a 1 percentage point increase in demand disturbance e_c are expressed as log deviations from steady state. σ_i/σ_c stands for the relative volatility of investment to consumption. Original series are Hodrick-Prescott (HP) filtered with a smoothing parameter of 1,600 to calculate the volatility. The U.S. data is from the BEA. A path of 5,000 quarters is simulated to calculate the statistics for each calibrated model.

5 Estimating the Capacity Underutilization Model

In this section, I extend the basic CU model to a full CU model. I estimate the full CU model using Bayesian estimation techniques. Based on the estimated results, I show that demand shocks are the main driving forces of business cycles. Particularly, a single consumption demand shock in the full CU model can already explain most of the business cycle fluctuations and co-movements.

[22]If investment adjustment cost is introduced to obtain a hump-shaped impulse response function of investment, the volatility of investment will decrease and it becomes even more difficult for us to achieve a realistic relative volatility of investment to consumption in the standard VU model.

[23]Although an increase in capital utilization increases output for a given amount of capital and labor, the cost of capital utilization also increases. These two have opposite effects on the Solow residual, and at a first order approximation, they cancel out.

5.1 Model Extensions

The full CU model is obtained by extending the basic CU model with home production, indirect labor, capital and investment adjustment costs, and exogenous expenditure.

5.1.1 Home Production

I assume that households use their time to produce goods or services at home. The total hours available is normalized to one. The amount of home produced goods or services is given by $c_{h,t} = Z_{h,t}(1 - l_t)$, where $Z_{h,t}$ is the productivity of time spent at home. The instantaneous utility function of the representative household is now given by

$$u(z_{c,t}, c_t, c_{h,t}) = \begin{cases} \phi e^{\gamma z_{c,t}} \frac{c_t^{1-\gamma}-1}{1-\gamma} + \bar{\omega} c_{h,t}, & \gamma \neq 1 \\ \phi e^{z_{c,t}} \ln(c_t) + \bar{\omega} c_{h,t}, & \gamma = 1 \end{cases} \tag{5.1}$$

which replaces the utility function (2.15) in the household's problem.

The inclusion of home production enriches the model dynamics. An increase in the productivity at home increases the opportunity cost of supplying labor, and thus reduces the household's desire to consume and the firm's desire to invest. Later in section 5.1, I will relate the productivity at home to labor productivity shock $z_{l,t}$: $Z_{h,t} = e^{\phi_l z_{l,t}}$. The parameter ϕ_l is usually set to be zero in standard models. In this case, an increase in labor productivity reduces the required labor to produce output but does *not* increase the opportunity cost of supplying labor. Hence, output moves positively with labor productivity shock. However, if ϕ_l is positive, an increase in labor productivity also gives an increase in the productivity at home. If ϕ_l is equal to one, the two opposite forces on output cancel out and the increase in labor productivity has to be absorbed by a decrease in working hours. Hence, the magnitude of ϕ_l affects the importance of labor productivity in driving business cycles. I will estimate ϕ_l later in section 5.2.

5.1.2 Indirect Labor

I also assume that the Leontief production function of firm $j \in [0,1]$ at time t is now given by

$$y_{j,t} = \min\left\{ A k_{j,t}, \frac{l_{f,j,t}}{\alpha_{f,t}}, \frac{l_{v,j,t}}{\alpha_{v,t}} \right\}, \tag{5.2}$$

where $l_{v,j,t}$ is the amount of direct labor, $l_{f,j,t}$ is the amount of indirect labor, and the inverse of $\alpha_{f,t}$ is the productivity of the indirect labor.

Direct labor produces goods or services directly. The hours of direct labor fluctuate with output and can be easily adjusted within a period. Examples of direct labor positions are machine operators, assembly line operators, and cleaners.

Indirect labor, by contrast, supports the production process of the firm, but is not directly involved in the active conversion of materials into finished products. Like capital stock, indirect labor is predetermined. Examples of indirect labor positions are production supervisor, managerial and various administrative labor positions, such as accounting, marketing, and human resource positions.

By definition, the production capacity of firm j at time t is determined by the short run fixed factors, i.e., indirect labor and capital stock:

$$\bar{y}_{j,t} = \min \left\{ Ak_{j,t}, \frac{l_{f,j,t}}{\alpha_{f,t}} \right\}, \tag{5.3}$$

where $\alpha_{f,t}$ can also be understood as the amount of indirect labor required to form a unit of capacity.

The introduction of indirect labor allows output to be more volatile than labor. Hence, labor productivity measured by output to labor ratio can be pro-cyclical under demand shocks.

5.1.3 Capital and Investment Adjustment Costs

To get a persistent and hump-shaped investment response, I introduce capital and investment adjustment costs. The capital stock of firm j is accumulated according to

$$k_{j,t+1} = k_{j,t}(1-\delta) + i_{j,t}(1 - S(i_{j,t}, i_{j,t-1}, k_{j,t})), \tag{5.4}$$

where S is the adjustment cost function.

Following Hayashi (1982) and Christiano et al. (2005), I assume that the adjustment cost function is given by

$$S(i_{j,t}, i_{j,t-1}, k_{j,t}) = \frac{\phi_i}{2}\left(\frac{i_{j,t}}{i_{j,t-1}} - 1\right)^2 + \frac{\phi_k}{2}\left(\frac{i_{j,t}}{k_{j,t}} - \delta\right)^2 \frac{k_{j,t}}{i_{j,t}}, \tag{5.5}$$

where $\phi_k \geq 0$ and $\phi_i \geq 0$ are parameters that capture the curvature of the capital and investment adjustment costs respectively.

This functional form implies that to change the level of investment or to deviate from the level of investment that maintains the current level of capital stock is costly. Hence, the adjustment costs are zero in steady state, but the dynamics around the steady state will be influenced by the curvature of these two adjustment cost components.

5.1.4 Exogenous Expenditure

Following Smets and Wouters (2007), government spending and net exports are treated as exogenous expenditure. I abstract away from the crowding-out and (or) crowding-in effects of government spending on consumption and investment by assuming that the exogenous expenditure is produced by an independent sector that requires direct labor only. Let g_t be the exogenous expenditure and $l_{g,t}$ be the corresponding labor hired. We have $g_t = l_{g,t}/\alpha_{g,t}$, where $\alpha_{g,t}$ is the direct labor required per exogenous expenditure. The final product of the whole economy is given by

$$Y_t = y_t + g_t, \tag{5.6}$$

where $y_t = c_t + i_t$ is the output of the private sector, and the total hours worked is given by

$$l_t = \alpha_{f,t} A k_t + \alpha_{v,t} y_t + \alpha_{g,t} g_t. \tag{5.7}$$

5.1.5 Exogenous Shocks

I introduce four exogenous shocks in total.

First, an exogenous expenditure shock $z_{g,t}$ drives directly the exogenous expenditure: $g_t = g e^{z_{g,t}}$. I assume that $z_{g,t}$ follows an AR(1) process with an i.i.d. Normal error term: $z_{g,t} = \rho_g z_{g,t-1} + e_{g,t}$.

Second, a labor productivity shock $z_{l,t}$ is assumed to follow an AR(1) process with an i.i.d. Normal error term: $z_{l,t} = \rho_l z_{l,t-1} + e_{l,t}$; and $z_{l,t}$ determines the productivity of both types of labor. Specifically, $\alpha_{v,t} = \alpha_v e^{-z_{l,t}}$, $\alpha_{f,t} = \alpha_f e^{-z_{l,t}}$, and $\alpha_{g,t} = \alpha_g e^{-z_{l,t}}$. In addition, the productivity of time spent at home is also affected by the labor productivity shock: $Z_{h,t} = e^{\phi_l z_{l,t}}$.

Third, an investment demand shock $z_{i,t}$ affects the subjective discount factor, i.e., the importance of the present relative to the future. The lifetime preference of the representative household is now given by:

$$\mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t e^{-z_{i,t}} u\left(z_{c,t}, c_t, c_{h,t}\right) \right). \tag{5.8}$$

I assume that $z_{i,t}$ follows an AR(2) processes with an i.i.d. Normal error term: $z_{i,t} = \rho_{i,1} z_{i,t-1} + \rho_{i,2}\left(z_{i,t-1} - z_{i,t-2}\right) + e_{i,t}$. The assumption is designed to capture both the low-frequency movements and the high-frequency zigzag pattern in the investment demand.

Finally, as in the basic CU model, a consumption demand shock $z_{c,t}$ affects the marginal utility of consumption relative to the marginal dis-utility of labor. I assume that $z_{c,t}$ follows an AR(1) process with an i.i.d. Normal error term and is also affected by the error term of the investment demand shock as follows:

$$z_{c,t} = \rho_c z_{c,t-1} + e_{c,t} + \rho_{ci} e_{i,t}.$$

The parameter ρ_{ci} turns out to capture a slow movement in consumption that is negatively related to investment but is not much related to business cycles.[24]

All error terms, $e_{c,t}$, $e_{i,t}$, $e_{g,t}$, and $e_{l,t}$, are mutually uncorrelated.

5.2 Estimation

I use the Bayesian estimation techniques to estimate the full CU model.[25] Having four exogenous shocks, I am able to match four detrended U.S. macro time series: real consumption, real investment, real exogenous expenditure, and the amount of hours worked.[26]

To alleviate the burden on estimation, I choose nine calibration targets that must be matched throughout this estimation procedure. Eight of them are the same as described in section 3.5. One target is new, which is the exogenous expenditure to output ratio in steady state.

Because of these targets, nine parameter values are *not* free to pick. Six parameters are fixed as their values follow directly from the nine calibration targets. Three parameters are expressed as a *function* of the other parameters and the nine calibration targets:

$$\beta = \left((1 - \bar{\omega}\alpha_v) Au - \bar{\omega}\alpha_f A + 1 - \delta \right)^{-1}, \tag{5.9}$$

$$\alpha_f = (l - \alpha_v (c + i) - \alpha_g g) \frac{\delta}{Ai}, \tag{5.10}$$

and

$$\varepsilon_D = (1 - \bar{\omega}\alpha_v)^{-1}. \tag{5.11}$$

Table 4 summarizes these nine parameters and their mostly associated calibration targets.

The other parameters are estimated and their priors are assumed as follows. The standard errors of the innovations follow an inverse-gamma distribution with a mean of 0.03 and a standard deviation of infinity. The persistence parameters of the stochastic processes, ρ_c, ρ_g, $\rho_{i,1}$, and ρ_l, follow a uniform distribution which ranges from -1 to 1.

[24] Appendix E gives some further discussions and shows the estimated consumption and investment series under the investment demand shock $e_{i,t}$.

[25] The model is log-linearized around steady state and the estimation procedure is done with the software platform Dynare.

[26] See Appendix B for a full description of the data.

Table 4: Parameters Pinned down by the Calibration Targets – Full CU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	2.0000	Inter-temporal elasticity of substitution 0.5
A	0.1544	Investment to output ratio 0.17
g	0.2000	Exogenous expenditure to output ratio 0.2
$\bar{\omega}$	0.6798	Output normalized to 1
ϕ	0.3969	Marginal utility of income normalized to 1
β		Capacity utilization rate 0.8
α_f		Labor underutilization rate 0.088
ε_D		Labor share of income 0.62

The parameter $\rho_{i,2}$ that allows the investment demand to have a zigzag pattern and the parameter ρ_{ci} that allows the investment innovation to have an effect on the consumption demand both follow a Normal distribution with a mean of 0 and a standard deviation of 0.2. The parameter ϕ_l that gives the relationship between the productivity at home and the productivity of labor is uniformly distributed within the range from -1 to 1. The curvature of the capital adjustment cost ϕ_k is Normal distributed around 2 with a standard deviation of 1. The curvature of the investment adjustment cost ϕ_i is Normal distributed around 0.2 with a standard deviation of 0.1. The direct labor required per output in private sector α_v and the direct labor required per exogenous expenditure α_g are both Normal distributed around 0.6 with a standard error of 0.3.

Table 5 summarizes the priors and shows the mode, the mean, and the 5th and 95th percentiles of the posterior distribution of the parameters obtained by the Metropolis-Hastings (MH) algorithm.[27]

Based on the posterior modes of the structural parameters and the nine calibration targets, the subjective discount factor β is 0.99, the indirect labor required per unit of capacity α_f is 0.20, and the elasticity of the demand curve ε_D is 2.06.

5.3 Variance Decomposition

What are the main driving forces of business cycles?

Table 6 gives the forecast error variance decomposition of output, consumption, investment, hours, and the capacity utilization rate at a 10-year horizon based on the posterior modes of the parameters of the full CU model.

According to the estimated full CU model, business cycle movements are primarily driven by three types of demand shocks, i.e., the consumption demand, the investment

[27] Total number of MH draws is 100,000 and the acceptance ratio is about 23.5%.

Table 5: Bayesian Estimation – Full CU Model

	Prior Distribution			Posterior Distribution			
Parameter	Distribution	Mean	Std Dev	Mode	Mean	5th Percentile	95th Percentile
σ_c	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
σ_i	Invgamma	0.03	∞	0.03	0.03	0.02	0.05
σ_g	Invgamma	0.03	∞	0.02	0.02	0.02	0.03
σ_l	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
ρ_c	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
$\rho_{i,1}$	Uniform	0.00	0.58	0.99	0.98	0.97	0.99
$\rho_{i,2}$	Normal	0.00	0.20	-0.25	-0.28	-0.38	-0.17
ρ_g	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
ρ_l	Uniform	0.00	0.58	0.96	0.96	0.93	0.99
ρ_{ci}	Normal	0.00	0.20	-0.14	-0.14	-0.19	-0.07
ϕ_l	Uniform	0.00	0.58	0.64	0.64	0.39	0.92
ϕ_k	Normal	2.00	1.00	2.01	2.17	1.20	3.07
ϕ_i	Normal	0.20	0.10	0.18	0.19	0.128	0.26
α_v	Normal	0.60	0.30	0.76	0.75	0.63	0.86
α_g	Normal	0.60	0.30	0.55	0.55	0.42	0.67

Note: Std Dev stands for standard deviation (of the priors). The sample period is from the first quarter of 1948 to the first quarter of 2017.

Table 6: Forecast Error Variance Decomposition (%)

		Full CU Model				Full VU Model			
		e_c	e_i	e_g	e_l	e_c	e_i	e_g	e_l
Output	Y	56.52	15.01	26.25	2.23	33.46	4.27	22.15	40.12
Consumption	c	72.65	26.45	0.00	0.01	64.89	15.89	0.00	19.23
Investment	i	78.70	12.75	0.00	8.56	12.53	0.74	0.00	86.73
Hours	l	60.84	14.80	10.84	13.42	66.67	11.54	11.38	10.41
CU Rate	u	71.50	22.03	0.00	6.47	27.15	9.63	0.00	67.62

Note: The forecast error variance decomposition for the estimated models at a 10-year (40-quarter) horizon. CU Rate stands for capacity utilization rate.

demand, and the exogenous expenditure shocks. The labor productivity shock accounts for only 2.23 percent of the variation in output and 13.42 percent of the variation in hours. Among these three types of demand shocks, the consumption demand shock turns out to be the main driver of business cycles. The single consumption demand shock can already explain most of the variations in consumption, investment, hours, and the capacity utilization rate.

For comparison, I extend the standard VU model described in section 4.1 by including home production, capital and investment adjustment costs, exogenous expenditure, and the four exogenous shocks in the same way as I did for the basic CU model. The extended VU model is called the full VU model. The calibration and estimation procedure of the full VU model is set to be as close as possible to that of the full CU model.[28] The variance decomposition results of the full VU model are listed in Table 6.

Consistent with the standard business cycle literature, the primary driving force of output in the full VU model is the labor productivity shock, which explains most of the variations in investment. However, the labor productivity shock is unable to explain most of the variations in hours (see Smets and Wouters, 2007, for a related discussion). In addition, consumption is not volatile enough under the labor productivity shock (see Bai et al., 2012, for similar results found in a standard RBC model). Therefore, standard models often rely on multiple types of shocks to explain business cycles.

5.4 Impulse Responses and Discussions

To understand the variance decomposition results better, I plot the IRFs of both the full CU model and the full VU model in Figure 6 based on the parameter values evaluated at the posterior modes.

In the full CU model, a positive innovation to consumption demand leads to a large increase in consumption, investment, output, and hours. Not surprisingly, the consumption demand innovations are the main driving forces of business cycles. A positive innovation to investment demand leads to an increase in investment but a small decrease in consumption because the parameter $\rho_{ci} = -0.14$ is estimated to be negative. A positive innovation to exogenous expenditure increases output and hours but has no effect on private consumption and investment because the exogenous expenditure is assumed to be produced independently from the private sector. Finally, since ϕ_l is estimated to be about 0.64, which is quite close to 1, a positive innovation to labor productivity has only a small effect on consumption and investment. Hence, output does not increase that much and the amount of hours decreases as labor is displaced by the improvement in

[28]See Appendix D for a detailed description of the calibration and estimation of the full VU model.

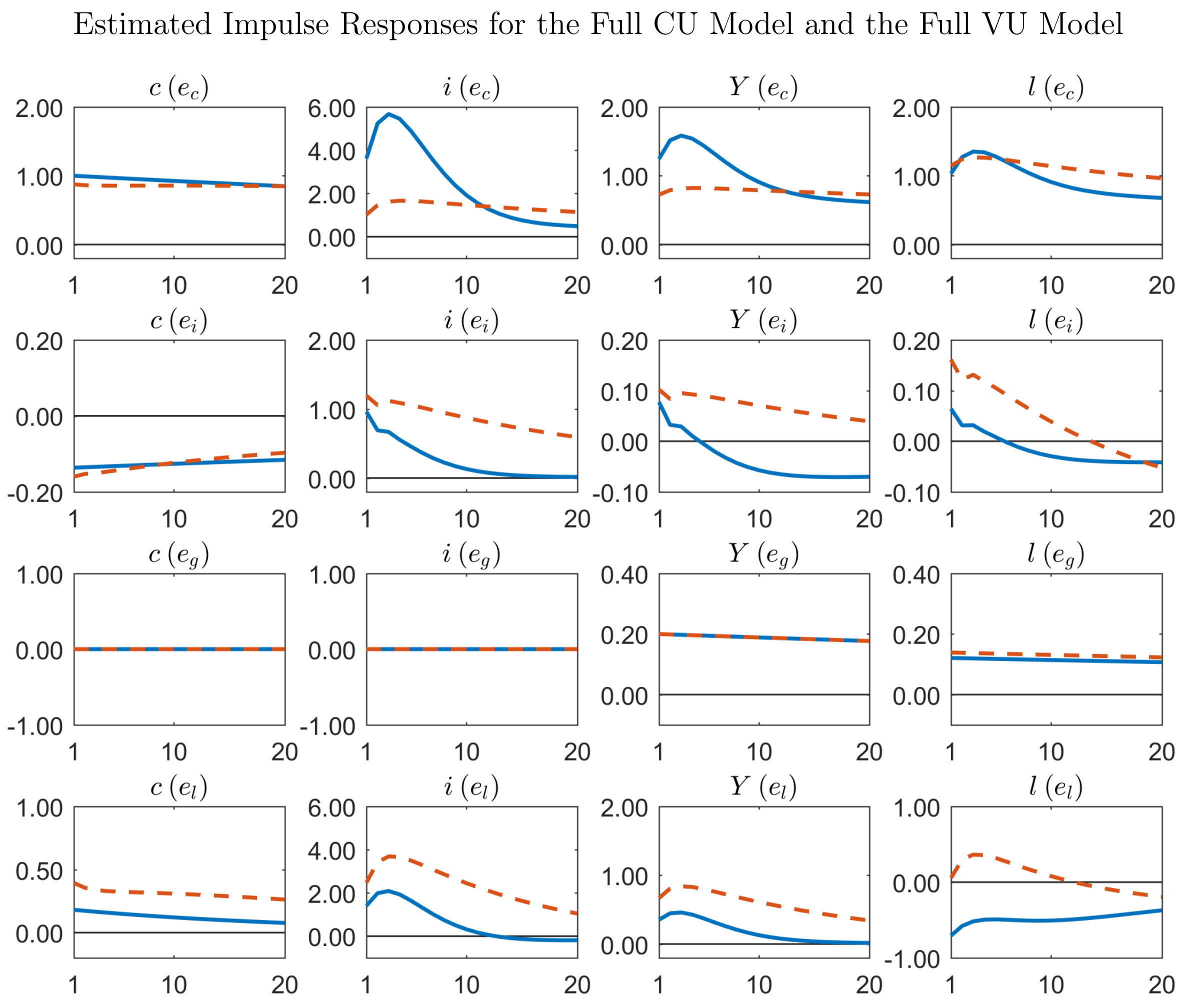

Figure 6: Responses to a 1 percentage point increase in disturbances e_c, e_i, e_g, and e_l. The solid lines are the impulse responses for the full CU model. The dashed lines are the impulse responses for the full VU model. All variables are expressed as log deviations from steady state.

labor productivity.

In the full VU model, a positive innovation to consumption demand generates a sizable increase in consumption and hours but fails to generate a large increase in investment. A positive innovation to investment demand or to exogenous expenditure does not generate the desired business cycle co-movements as in the full CU model. Finally, since ϕ_l is estimated to be about -0.02 in the full VU model, which is very close to 0, a positive innovation to labor productivity can have a large effect on output. However, as labor is displaced by the improvement in labor productivity, hours fail to increase much. Hence, no *single* shock in the full VU model is able to drive the business cycles.

5.5 Consumption Demand Driven Business Cycles

To highlight the role of consumption demand shock in driving business cycles, I shut down all the other shocks and feed only the consumption demand shock into the models. The structural shocks of the estimated models can be backed out from the U.S. data using the Kalman smoother technique.

Model Predicted Series under the Consumption Demand Shock

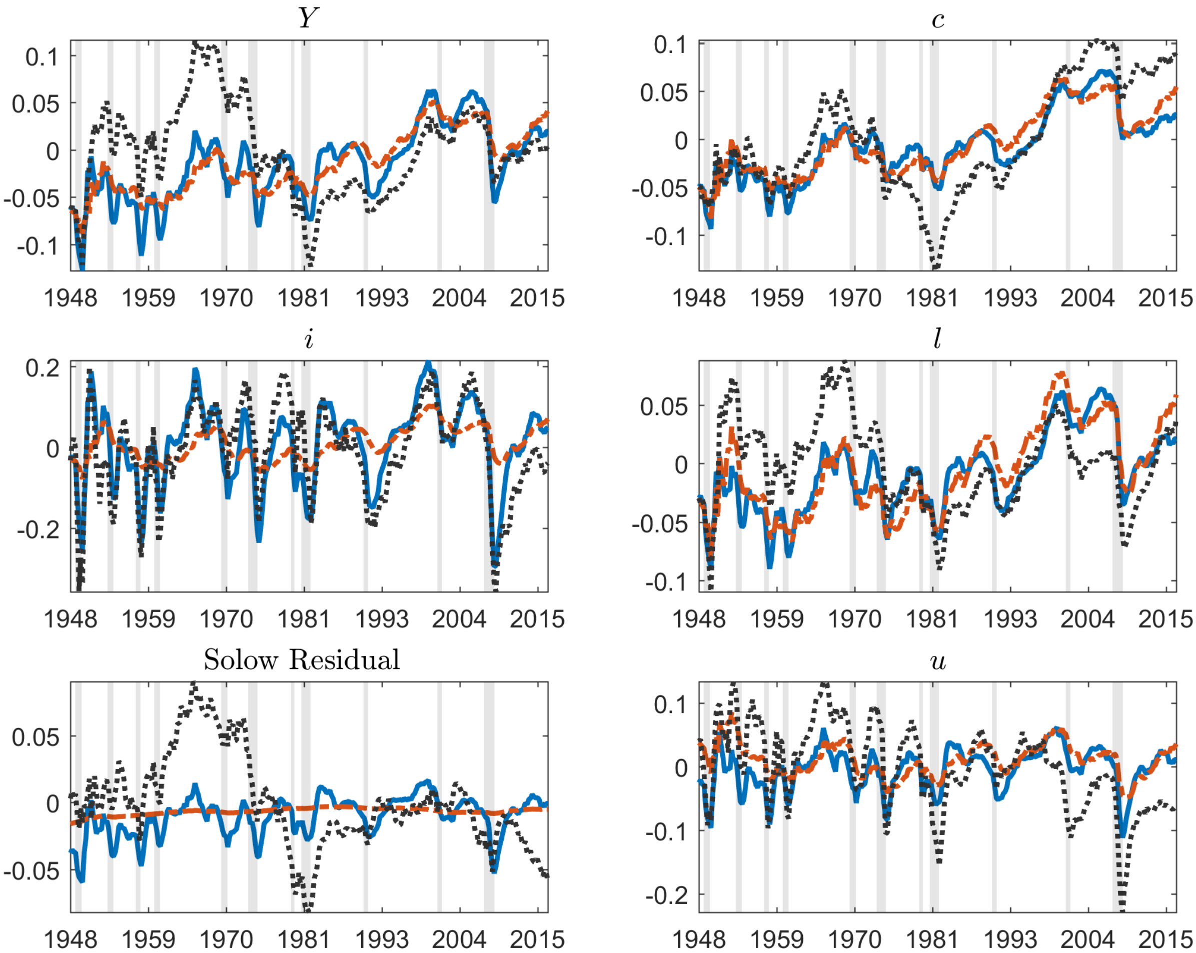

Figure 7: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Figure 7 shows the model predicted series under the single consumption demand shock. The series predicted by the full CU model fit the U.S. data well. In the full VU model, however, investment has little response to the consumption demand shock and the Solow residual almost does not fluctuate.

Table 7 compares the main business cycle statistics of the predicted series with those of the U.S. data. The business cycle properties of the full CU model under the consump-

Table 7: Business Cycle Statistics

	The U.S. Data		Full CU Model (e_c)		Full VU Model (e_c)		Full VU Model (e_l)	
	Std Dev	Cov w Y	Std Dev	Cov w Y	Std Dev	Cov w Y	Std Dev	Cov w Y
Output	1.62	1.00	1.79	1.00	0.78	1.00	1.58	1.00
Consumption	1.16	0.57	1.03	0.55	0.82	1.06	0.64	0.39
Investment	7.55	3.95	6.96	3.83	1.56	1.97	7.02	4.42
Hours	1.68	0.89	1.45	0.81	1.23	1.58	0.81	0.47
Solow Resid	1.00	0.39	0.93	0.50	0.03	0.02	1.17	0.72
Real Wage Rate	0.97	-0.16	0.00	0.00	0.27	-0.35	1.26	0.77

Note: Std Dev stands for standard deviation. Cov w Y stands for covariance with output. Solow Resid stands for Solow residual. Covariance with output is reported relative to the variance of output. The U.S. data is from the BEA and the BLS. All variables are HP-filtered logarithms of the original series.

tion demand shock alone are already quite close to those of the data. The consumption demand shock not only generates sizable business cycle fluctuations but also generates the correct business cycle co-movements between the key aggregate variables. Particularly, investment is highly volatile and the Solow residual is as pro-cyclical as in the data. Because of the capital resource slackness, the real wage rate in the full CU model is independent of the size of the demand and is fully determined by the labor productivity. Hence, the real wage rate is naturally acyclical in the full CU model under the consumption demand shock, a feature that is also close to that observed in the data.

I also calculate the main business cycle statistics for the series predicted by the full VU model under the labor productivity shock. The results are listed in the last three columns of Table 7. Although the labor productivity shock in the full VU model is able to generate large fluctuations in investment and the Solow residual, it fails to generate large movements in labor hours and consumption. In addition, the real wage rate in the full VU model under the labor productivity shock looks too pro-cyclical.

To sum up, in the full VU model, the consumption demand shock fails to create a large response in investment. Thus, the full VU model relies on labor productivity shock to obtain a volatile investment. Although the labor productivity shock is an important driving force of business cycle in standard models, it is not dominating because hours and consumption are typically not volatile enough under the single labor productivity shock. By contrast, when viewed through the lens of the full CU model, the consumption demand shock alone can already explain most of the business cycle fluctuations and co-movements.

6 Conclusions

I build a simple macroeconomic model that incorporates capacity underutilization explicitly. Buyers are not fully attentive to prices as they are subject to an information processing cost. Thus, they search for capacity in a somewhat blind and random way. Hence, firms with a larger capacity can attract more buyers. The capacity competition among firms leads to long-term capacity underutilization. Since output is below capacity, the capital resources are slack. As a result, the real wage rate is acyclical and output can be highly responsive to demand shocks. I estimate a version of the model that features shocks to consumption demand, investment demand, and exogenous expenditure, as well as a labor productivity shock. Quantitatively, the consumption demand shock alone already explains most of the business cycle fluctuations, generating a large response in investment and hours, an acyclical real wage rate, and a pro-cyclical Solow residual, whereas the labor productivity shock explains only a small fraction of the variations in output and hours. Hence, when viewed through the lens of my capacity underutilization model, demand is the main driving force of business cycles.

References

Anderson, E. W., Fornell, C., Lehmann, D. R. 1994. Customer satisfaction, market share, and profitability: Findings from sweden. Journal of marketing, 58, 53–66.

Andrle, M., Bruha, J., Solmaz, S. 2017. On the sources of business cycles: implications for DSGE models. ECB Working Paper.

Angeletos, G.-M., Collard, F., Dellas, H. 2018. Business cycle anatomy.Technical report, National Bureau of Economic Research.

Athanassopoulos, A. D. 2000. Customer satisfaction cues to support market segmentation and explain switching behavior. Journal of business research, 47, 191–207.

Bai, Y., Rios-Rull, J.-V., Storesletten, K. 2012. Demand shocks as productivity shocks. Federal Reserve Board of Minneapolis.

Basu, S., Kimball, M. S. 1997. Cyclical productivity with unobserved input variation.Technical report, National Bureau of Economic Research.

Bils, M., Cho, J.-O. 1994. Cyclical factor utilization. Journal of Monetary Economics, 33, 319–354.

Boehm, C. E., Flaaen, A., Pandalai-Nayar, N. 2017. Are supply curves convex? implications for state-dependent responses to shocks. Working paper.

Buch, C. M., Döpke, J., Strotmann, H. 2009. Does export openness increase firm-level output volatility? World Economy, 32, 531–551.

Bulow, J., Geanakoplos, J., Klemperer, P. 1985. Holding idle capacity to deter entry. The Economic Journal, 95, 178–182.

Burnside, C., Eichenbaum, M., Rebelo, S. 1993. Labor hoarding and the business cycle. Journal of Political Economy, 101, 245–273.

Christiano, L. J., Eichenbaum, M., Evans, C. L. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of political Economy, 113, 1–45.

Comin, D., Philippon, T. 2005. The rise in firm-level volatility: Causes and consequences. NBER macroeconomics annual, 20, 167–201.

Cooley, T. F., Hansen, G. D., Prescott, E. C. 1995. Equilibrium business cycles with idle resources and variable capacity utilization. Economic Theory, 6, 35–49.

Corrado, C., Mattey, J. 1997. Capacity utilization. The Journal of Economic Perspectives, 11, 151–167.

De Veirman, E., Levin, A. 2018. Cyclical changes in firm volatility. Journal of Money, Credit and Banking, 50, 317–349.

Deng, S., Yano, C. A. 2006. Joint production and pricing decisions with setup costs and capacity constraints. Management Science, 52, 741–756.

Dixit, A. 1980. The role of investment in entry-deterrence. The economic journal, 90, 95–106.

Eiteman, W. J. 1947. Factors determining the location of the least cost point. The American Economic Review, 37, 910–918.

Eiteman, W. J., Guthrie, G. E. 1952. The shape of the average cost curve. The American economic review, 42, 832–838.

Fagnart, J.-F., Licandro, O., Portier, F. 1999. Firm heterogeneity, capacity utilization, and the business cycle. Review of Economic Dynamics, 2, 433–455.

Fagnart, J.-F., Licandro, O., Sneessens, H. R. 1997. Capacity utilization and market power. Journal of Economic Dynamics and Control, 22, 123–140.

Florian, M., Klein, M. 1971. Deterministic production planning with concave costs and capacity constraints. Management Science, 18, 12–20.

Gilchrist, S., Williams, J. C. 2000. Putty-clay and investment: a business cycle analysis. Journal of Political Economy, 108, 928–960.

Greenwood, J., Hercowitz, Z., Huffman, G. W. 1988. Investment, capacity utilization, and the real business cycle. The American Economic Review, 402–417.

Gupta, S., Starr, M. 2014. Production and operations management systems. CRC Press.

Hall, R. E. 1986. Chronic excess capacity in US industry.

Hansen, G. D. 1985. Indivisible labor and the business cycle. Journal of monetary Economics, 16, 309–327.

Hayashi, F. 1982. Tobin's marginal q and average q: A neoclassical interpretation. Econometrica: Journal of the Econometric Society, 213–224.

Johansen, L. 1959. Substitution versus fixed production coefficients in the theory of economic growth: a synthesis. Econometrica: Journal of the Econometric Society, 157–176.

Kalish, S. 1983. Monopolist pricing with dynamic demand and production cost. Marketing Science, 2, 135–159.

Kelly, B., Lustig, H., Van Nieuwerburgh, S. 2013. Firm volatility in granular networks.Technical report, National Bureau of Economic Research.

Klein, L. R. 1960. Some theoretical issues in the measurement of capacity. Econometrica: Journal of the Econometric Society, 272–286.

Köberl, E. M., Lein, S. M. 2011. The nircu and the phillips curve: an approach based on micro data. Canadian Journal of Economics/Revue canadienne d'économique, 44, 673–694.

Kumar, S. A., Suresh, N. 2006. Production and operations management. New Age International.

Kumar, V., Batista, L., Maull, R. 2011. The impact of operations performance on customer loyalty. Service Science, 3, 158–171.

Kumar, V., Kumari, A., Garza-Reyes, J. A., Lim, M. 2013. Dependability a key element for achieving competitive advantage: A study of information service firms. In IFIP International Conference on Advances in Production Management Systems, 493–500, Springer.

Kydland, F. E., Prescott, E. C. 1988. The workweek of capital and its cyclical implications. Journal of Monetary Economics, 21, 343–360.

Matějka, F., McKay, A. 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. American Economic Review, 105, 272–98.

Mattsson, L.-G., Weibull, J. W. 2002. Probabilistic choice and procedurally bounded rationality. Games and Economic Behavior, 41, 61–78.

Michaillat, P., Saez, E. 2015. Aggregate demand, idle time, and unemployment. The Quarterly Journal of Economics, 130, 507–569.

Moen, E. R. 1997. Competitive search equilibrium. Journal of political Economy, 105, 385–411.

Morrison, C. J. 1985. Primal and dual capacity utilization: an application to productivity measurement in the us automobile industry. Journal of Business & Economic Statistics, 3, 312–324.

Nelson, R. A. 1989. On the measurement of capacity utilization. The Journal of Industrial Economics, 273–286.

Olhager, J., Johansson, P. 2012. Linking long-term capacity management for manufacturing and service operations. Journal of Engineering and Technology Management, 29, 22–33.

Olhager, J., Rudberg, M., Wikner, J. 2001. Long-term capacity management: Linking the perspectives from manufacturing strategy and sales and operations planning. International Journal of Production Economics, 69, 215–225.

Rogerson, R. 1988. Indivisible labor, lotteries and equilibrium. Journal of monetary Economics, 21, 3–16.

Rust, R. T., Zahorik, A. J. 1993. Customer satisfaction, customer retention, and market share. Journal of retailing, 69, 193–215.

Smets, F., Wouters, R. 2007. Shocks and frictions in us business cycles: A bayesian dsge approach. American economic review, 97, 586–606.

Spence, A. M. 1977. Entry, capacity, investment and oligopolistic pricing. The Bell Journal of Economics, 534–544.

Stevenson, W. J. 2002. Operations management. McGraw-Hill/Irwin Boston.

Walters, A. A. 1963. Production and cost functions: An econometric survey. Econometrica: Journal of the Econometric Society, 1–66.

Zhang, Q., Vonderembse, M. A., Lim, J.-S. 2003. Manufacturing flexibility: defining and analyzing relationships among competence, capability, and customer satisfaction. Journal of Operations Management, 21, 173–191.

Zhang, Q., Vonderembse, M. A., Lim, J.-S. 2005. Logistics flexibility and its impact on customer satisfaction. The International journal of logistics management, 16, 71–95.

A Calibrations of the Standard RBC Models

I do not use the observed average capacity utilization rate as a calibration target because capacity is fully utilized in the steady state of the standard RBC models. The other calibration targets are the same as in the basic CU model (see section 3.5).

Table 8 summarizes the calibrated parameter values of the Leontief-RBC model and their mostly associated targets.

Table 8: Parameters and Calibration Targets – Leontief-RBC Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Inter-temporal elasticity of substitution 1
ϕ	0.8300	Marginal utility of income 1
$\bar{\omega}$	0.6798	Output 1
β	0.9747	Labor share of income 0.62
A	0.1235	Investment to output ratio 0.17
α_v	0.9120	Labor underutilization rate 0.088

For the CD-RBC model, the capital productivity A is not identified because for each value of $A > 0$, there exists a value of α_v such that the total factor productivity (TFP) $\alpha_v^{-\alpha} A^{1-\alpha}$ is the same. To pin down the value of A, I choose A such that $y/(Ak)$ is normalized to 1. Table 9 summarizes the calibrated parameter values of the CD-RBC model and their mostly associated targets.

Table 9: Parameters and Calibration Targets – CD-RBC Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Inter-temporal elasticity of substitution 1
ϕ	0.8300	Marginal utility of income 1
$\bar{\omega}$	0.6798	Output 1
β	0.9747	Investment to output ratio 0.17
A	0.1235	$y/(Ak)$ normalized to 1
α_v	0.9120	Labor underutilization rate 0.088
α	0.6200	Labor share of income 0.62

B Description of the Data

The data on consumption, investment, government spending, exports, imports, output, hours, and capital is from the National Income and Product Accounts (NIPA) published by the U.S. Bureau of Economic Analysis (BEA). The data on employment, unemployment, labor force, and labor share is from the U.S. Bureau of Labor Statistics (BLS). The data on capacity utilization rate is from the Federal Reserve Board (FRB).

The FRB publishes a capacity utilization rate for manufacturing and a capacity utilization rate for total industries. The latter covers manufacturing, mining, and electric and gas utilities. These two capacity utilization rates are very close to each other. The manufacturing capacity utilization rate, which has a longer history, is used to compare with the model predicted series in Figure 7.

The BEA data on capital is annually. To get the quarterly data, I use linear interpolation to impute the BEA annual data on capital. Let K_Y be the capital stock at the end of year Y. The capital stock at the end of year Y and quarter Q is taken to be

$$K_{Y,Q} = \exp\left(\ln K_{Y-1} + \frac{Q}{4}\left(\ln K_Y - \ln K_{Y-1}\right)\right).$$

The BEA data on hours worked by full-time and part-time employees is also annually. To get the quarterly data, I impute the BEA annual data on hours based on the information provided by the BLS. The BLS issues data on hours worked in business sectors at a quarterly frequency. Let H_Y be the hours worked by full-time and part-time employees in year Y and $H^B_{Y,Q}$ be the hours worked in business sectors in year Y and quarter Q. The hours worked by full-time and part-time employees in year Y and quarter Q is taken to be

$$H_{Y,Q} = H_Y \frac{H^B_{Y,Q}}{H^B_{Y,1} + H^B_{Y,2} + H^B_{Y,3} + H^B_{Y,4}}.$$

Wage rate is inferred from output, hours, and labor share:

$$W_{Y,Q} = \frac{Y_{Y,Q}}{H_{Y,Q}} S^L_{Y,Q},$$

where $Y_{Y,Q}$ is the gross domestic product (GDP) and $S^L_{Y,Q}$ is the share of labor in year Y and quarter Q.

To convert the nominal variables into real ones, I divide the nominal variables by the GDP deflator obtained from the BEA.

All variables are detrended before they are used for estimations or to calculate business cycle statistics. I estimate a quadratic trend for the log of labor productivity measured by output to hours ratio. I also estimate a quadratic trend for the log of hours worked per capita.

Let $\bar{LP}_{Y,Q}$ be the trend of labor productivity, $\bar{h}_{Y,Q}$ be trend of hours worked per capita, and $\bar{L}_{Y,Q}$ be the labor force in year Y and quarter Q. The real wage rate is assumed to be of the same trend as the labor productivity. $\bar{h}_{Y,Q}\bar{L}_{Y,Q}$ captures the potential hours that could be worked in the economy, which is treated as the trend of hours worked by full-time and part-time employees. $\bar{h}_{Y,Q}\bar{L}_{Y,Q}\bar{LP}_{Y,Q}$ captures the potential output that could be produced in the economy, which is treated as the trend of consumption, investment, government spending, exports, imports, output, and capital.

Figure 8 shows the logarithms of the detrended consumption, investment, output, hours, capital, and real wage rate.

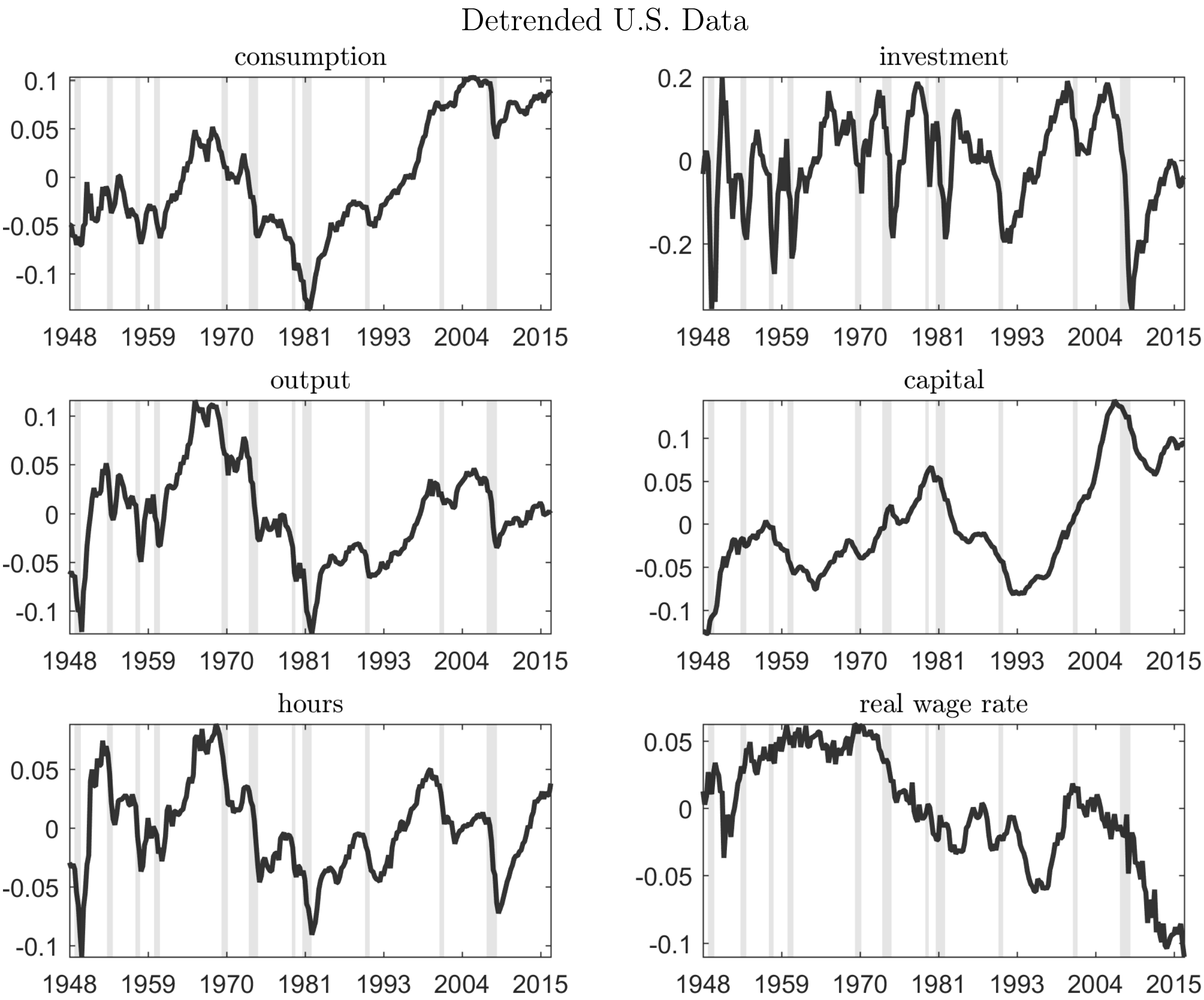

Figure 8: All variables are logarithms of the original series.

C Properties of the Variable Cost Function in the Standard VU Model

The following proposition gives the relevant properties of the variable cost function in the standard VU model.

Proposition 2. *At time t, the variable cost of firm j is defined by the following cost minimization problem:*

$$G(y_{j,t}, k_{j,t}, w_t) = \min_{l_{j,t}, \theta_{j,t}} w_t l_{j,t} + a(\theta_{j,t}) k_{j,t},$$

subject to the production constraint $F(l_{j,t}, \theta_{j,t} k_{j,t}) \geq y_{j,t}$. *Suppose that there is a unique interior solution to the above cost minimization problem.*

If $a(\cdot)$ *and* $F(\cdot, \cdot)$ *are twice differentiable,* $a(\cdot)$ *is strictly increasing and strictly convex,* $F(\cdot, \cdot)$ *is strictly increasing in both of its arguments, strictly concave, and homogeneous of degree one, we have that* $G(\cdot, \cdot, \cdot)$ *is twice differentiable, strictly convex, and homogeneous of degree one in terms of its first two arguments, i.e., output and capital. In addition,* $G_{yk} < 0$, *implying that capital expansion reduces the real marginal cost.*

Proof. Let $\lambda_{G,j,t}$ be the Lagrangian multiplier for the production constraint. The solution to the cost minimization is characterized by the following FOCs:

$$w_t = \lambda_{G,j,t} F_L(l_{j,t}, \theta_{j,t} k_{j,t}), \tag{C.1}$$

$$a'(\theta_{j,t}) = \lambda_{G,j,t} F_K(l_{j,t}, \theta_{j,t} k_{j,t}), \tag{C.2}$$

$$F(l_{j,t}, \theta_{j,t} k_{j,t}) = y_{j,t}. \tag{C.3}$$

Since $a(\theta_{j,t})$ and $F(l_{j,t}, \theta_{j,t} k_{j,t})$ are twice differentiable, we can total differentiate the above three equations at an arbitrary point where all variables are strictly positive:

$$B_{j,t} \begin{pmatrix} d\lambda_{G,j,t} \\ dl_{j,t} \\ d\theta_{j,t} \end{pmatrix} = \begin{pmatrix} \theta_{j,t}\lambda_{G,j,t}F_{LK,j,t} & 0 & -1 \\ \theta_{j,t}\lambda_{G,j,t}F_{KK,j,t} & 0 & 0 \\ -\theta_{j,t}F_{K,j,t} & 1 & 0 \end{pmatrix} \begin{pmatrix} dk_{j,t} \\ dy_{j,t} \\ dw_t \end{pmatrix},$$

in which matrix $B_{j,t}$ is given by

$$B_{j,t} = \begin{pmatrix} -F_{L,j,t} & -\lambda_{G,j,t}F_{LL,j,t} & -k_{j,t}\lambda_{G,j,t}F_{LK,j,t} \\ -F_{K,j,t} & -\lambda_{G,j,t}F_{KL,j,t} & \left(a''_{j,t} - k_{j,t}\lambda_{G,j,t}F_{KK,j,t}\right) \\ 0 & F_{L,j,t} & k_{j,t}F_{K,j,t} \end{pmatrix}.$$

Since $F_{L,j,t} > 0$, $F_{K,j,t} > 0$, $F_{LL,j,t} < 0$, $F_{KK,j,t} < 0$, $F_{KL,j,t} > 0$, and $a''_{j,t} > 0$, one can easily verify that the determinant of B is strictly positive.

Then, it follows from the implicit function theorem that $\lambda^*_{G,j,t}$, $l^*_{j,t}$, and $\theta^*_{j,t}$ that solve the minimization problem are differentiable functions of $k_{j,t}$, $y_{j,t}$, and w_t: $\lambda^*_{G,j,t} = \lambda^*_G(y_{j,t}, k_{j,t}, w_t)$, $l^*_{j,t} = l^*(y_{j,t}, k_{j,t}, w_t)$, $\theta^*_{j,t} = \theta^*(y_{j,t}, k_{j,t}, w_t)$. Hence, the variable cost function $G(y_{j,t}, k_{j,t}, w_t) = w_t l^*_{j,t} + a\left(\theta^*_{j,t}\right) k_{j,t}$ is differentiable.

According to the envelope theorem, $G_y(y_{j,t}, k_{j,t}, w_t) = \lambda^*_{G,j,t}$ and $G_k(y_{j,t}, k_{j,t}, w_t) = a(\theta^*_{j,t}) - \lambda^*_{G,j,t} F_K(l^*_{j,t}, \theta^*_{j,t} k_{j,t}) \theta^*_{j,t}$. Thus, G is twice differentiable.

In addition, by inverting B, we can get

$$G_{yk} = -\frac{\theta_{j,t}\lambda_{G,j,t}}{|B_{j,t}|}\left(\lambda_{G,j,t} k_{j,t} F_{K,j,t}\left(F^2_{KL,j,t} - F_{KK,j,t}F_{LL,j,t}\right) + a''_{j,t} F_{L,j,t} F_{LK,j,t}\right) < 0,$$

where $F^2_{KL,j,t} - F_{KK,j,t}F_{LL,j,t} > 0$ because $F(\cdot,\cdot)$ is strictly concave.

I now show that G is strictly convex in terms of output and capital. Suppose that θ_1 and l_1 minimizes the cost at (y_1, k_1, w); and θ_2 and l_2 minimizes the cost at (y_2, k_2, w). For any value $\nu \in (0,1)$, let $\tilde{l} = \nu l_1 + (1-\nu) l_2$, $\tilde{y} = \nu y_1 + (1-\nu) y_2$, $\tilde{k} = \nu k_1 + (1-\nu) k_2$, and $\varphi = \frac{\nu k_1}{\nu k_1 + (1-\nu) k_2} \in (0,1)$. We have:

$$\begin{aligned} F\left(\tilde{l}, (\varphi\theta_1 + (1-\varphi)\theta_2)\tilde{k}\right) &= F(\nu l_1 + (1-\nu) l_2, \nu\theta_1 k_1 + (1-\nu)\theta_2 k_2) \\ &> \nu F(l_1, \theta_1 k_1) + (1-\nu) F(l_2, \theta_2 k_2) \geq \tilde{y}. \end{aligned}$$

That is $(\varphi\theta_1 + (1-\varphi)\theta_2)$ and $\tilde{l}$ satisfy the production constraint at $\left(\tilde{y}, \tilde{k}, w\right)$. Therefore, we have:

$$\begin{aligned} \nu G(y_1, k_1, w) + (1-\nu) G(y_2, k_2, w) &= w\tilde{l} + (\varphi a(\theta_1) + (1-\varphi) a(\theta_2))\tilde{k} \\ &> w\tilde{l} + a(\varphi\theta_1 + (1-\varphi)\theta_2)\tilde{k}. \\ &\geq G\left(\tilde{y}, \tilde{k}, w\right). \end{aligned}$$

Hence, G is strictly convex.

Finally, I show that G is homogeneous of degree one in terms of output and capital. Suppose θ and l minimize the cost at (y, k, w). For any value $\nu > 0$, it is straight forward to verify that θ and νl minimize the cost at $(\nu y, \nu k, w)$. Hence, $G(\nu y, \nu k, w) = w\nu l + a(\theta)\nu k = \nu G(y, k, w)$. □

D Estimation of the Full VU Model

I use the Bayesian estimation techniques to estimate the full VU model. Thirteen calibration targets are chosen to be matched throughout this estimation procedure. Eleven of them are the same as described in section 4.3, which calibrates the standard VU model. Two targets are new. One is the exogenous expenditure to output ratio in steady state and the other is the real interest rate in steady state.

Because of these targets, thirteen parameter values are *not* free to pick. Eleven parameters are fixed as their values follow directly from the thirteen calibration targets.

Table 10: Parameters Pinned down by the Calibration Targets – Full VU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	2.0000	Inter-temporal elasticity of substitution 0.5
β	0.9900	Quarterly real interest rate 0.01
g	0.2000	Exogenous expenditure to output ratio 0.2
α_v	0.9120	Labor underutilization rate 0.088
$\bar{\omega}$	0.6798	Output normalized 1
ϕ	0.3969	Marginal utility of income normalized to 1
A	0.1235	$y/(Ak)$ normalized to 1
θ	1.0000	θ normalized to 1
$a(\theta)$	0.0000	$a(\theta)$ normalized to 0
α		Labor share of income 0.62
ε_D		Investment to output ratio 0.17

Two parameters are expressed as a *function* of the other parameters and the thirteen calibration targets:

$$\alpha = \frac{\bar{\omega}(l - \alpha_g g)}{\bar{\omega}(l - \alpha_g g) + (\beta^{-1} - 1 + \delta)\frac{i}{\delta}}, \tag{D.1}$$

and

$$\varepsilon_D = \frac{c + i}{c + i - \left(\bar{\omega}(l - \alpha_g g) + (\beta^{-1} - 1 + \delta)\frac{i}{\delta}\right)}. \tag{D.2}$$

Table 10 summarizes these thirteen parameters and their mostly associated calibration targets.

The other parameters are estimated. The priors of parameters σ_c, σ_i, σ_g, σ_l, ρ_c, ρ_g, $\rho_{i,1}$, $\rho_{i,2}$, ρ_l, ρ_{ci}, ϕ_l, ϕ_k, ϕ_i, and α_g are assumed to be the same as in the full CU model. The parameter $\xi_\theta^a \equiv a''(\theta)\theta/a'(\theta)$, which captures the convexity of the capital utilization cost function in steady state, is assumed to follow a Normal distribution with a mean of 1 and a standard deviation of 0.5.

Table 11 summarizes the priors and shows the mode, the mean, and the 5th and 95th percentiles of the posterior distribution of the parameters obtained by the Metropolis-Hastings (MH) algorithm. The total number of MH draws is 100,000 and the acceptance ratio is about 20.5%.

Based on the posterior modes of the structural parameters and the thirteen calibration targets, the Cobb-Douglas (CD) labor share α is 0.68 and the elasticity of the demand curve ε_D is 55.76.

Table 11: Bayesian Estimation – Full VU Model

	Prior Distribution			Posterior Distribution			
Parameter	Distribution	Mean	Std Dev	Mode	Mean	5th Percentile	95th Percentile
σ_c	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
σ_i	Invgamma	0.03	∞	0.03	0.02	0.02	0.03
σ_g	Invgamma	0.03	∞	0.02	0.02	0.02	0.03
σ_l	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
ρ_c	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
$\rho_{i,1}$	Uniform	0.00	0.58	0.98	0.98	0.97	0.99
$\rho_{i,2}$	Normal	0.00	0.20	-0.23	-0.25	-0.35	-0.16
ρ_g	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
ρ_l	Uniform	0.00	0.58	0.96	0.95	0.93	0.97
ρ_{ci}	Normal	0.00	0.20	-0.14	-0.15	-0.21	-0.08
ϕ_l	Uniform	0.00	0.58	-0.02	-0.03	-0.18	0.12
ϕ_k	Normal	2.00	1.00	1.88	2.09	0.85	3.29
ϕ_i	Normal	0.20	0.10	0.15	0.16	0.01	0.21
α_g	Normal	0.60	0.30	0.63	0.64	0.53	0.76
ξ_θ^a	Normal	1.00	0.50	0.97	1.11	0.37	1.79

Note: Std Dev stands for standard deviation (of the priors). The sample period is from the first quarter of 1948 to the first quarter of 2017.

E Additional Estimation Results

Figure 9 shows the consumption demand $z_{c,t}$, the investment demand $z_{i,t}$, the exogenous expenditure $z_{g,t}$, and the labor productivity $z_{l,t}$ series extracted from the data. Except for the labor productivity, the other exogenous stochastic processes extracted from the data are quite similar in both models.

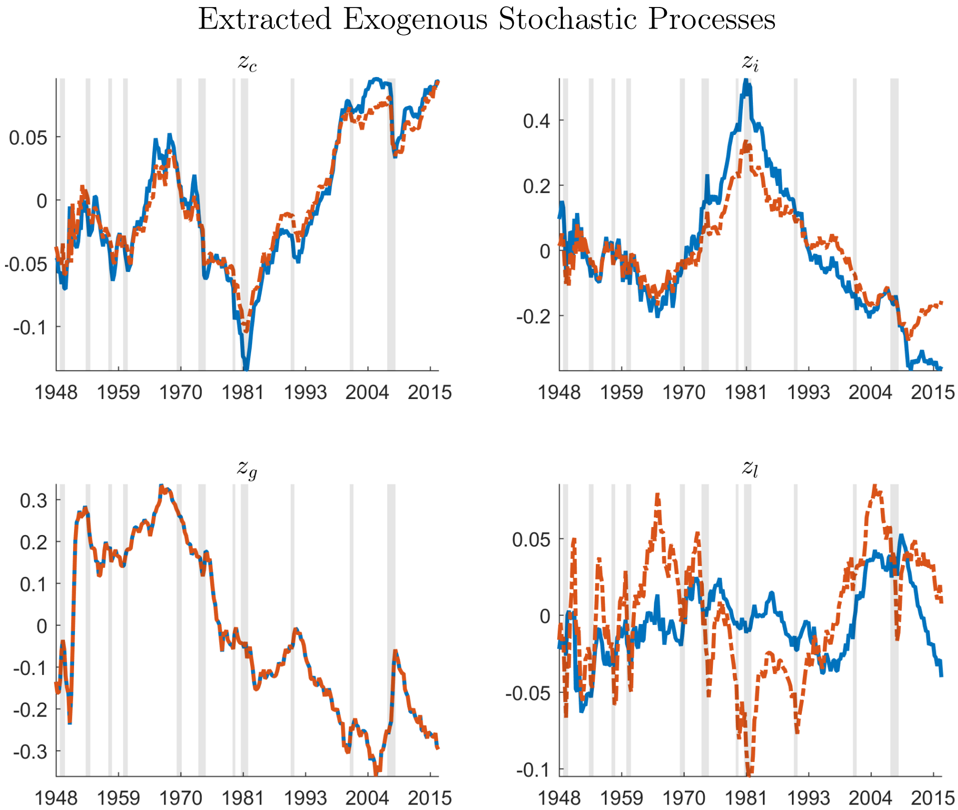

Figure 9: The consumption demand $z_{c,t}$, the investment demand $z_{i,t}$, the exogenous expenditure $z_{g,t}$, and the labor productivity $z_{l,t}$ extracted from the data.

Figure 10 shows the model predicted series under the single investment demand shock $e_{i,t}$. Since the estimated ρ_{ci} is negative, there is a negative correlation between the consumption and the investment. According to both models, investment shall be made to adjust capital to fit the long run movement in output (or consumption). However, in the U.S. data, despite big long-term swings in the detrended output (or consumption), detrended investment is fairly stable. The mismatch between the output (or consumption) level and the investment level is also reflected by the mismatch between the output (or consumption) level and the capital level (see Figure 8). This mismatch is unable to be captured by the internal mechanisms of both models but is captured by the parameter ρ_{ci} that allows the exogenous consumption demand and investment demand to have a negative correlation instead. The shock $e_{i,t}$, however, does not drive business cycles as consumption and investment move in opposite directions under this shock. In addition, the consumption movements under the investment demand shock do not exhibit large declines during the NBER dated recessions but rather capture the slow swings in the level of consumption.

Figure 11 shows the model predicted series under the single exogenous expenditure shock $e_{g,t}$. By construction, $e_{g,t}$ has no effect on consumption, investment, and capacity utilization rate.

Figure 12 shows the model predicted series under the single labor productivity shock $e_{l,t}$. The labor productivity shock plays a small role in the full CU model but an important role in the full VU model to drive cyclical movements in investment, the Solow residual, and the capacity utilization rate. The labor productivity shock, however, does not explain the fluctuations of hours very well and consumption is a bit too smooth under the labor productivity shock in the full VU model.

Figure 13 shows the model predicted real wage rate, which is not targeted in estimation procedures. The overall trend of the real wage rate in both models is roughly consistent with that of the data. However, the real wage rate is a bit too volatile in the full VU model.

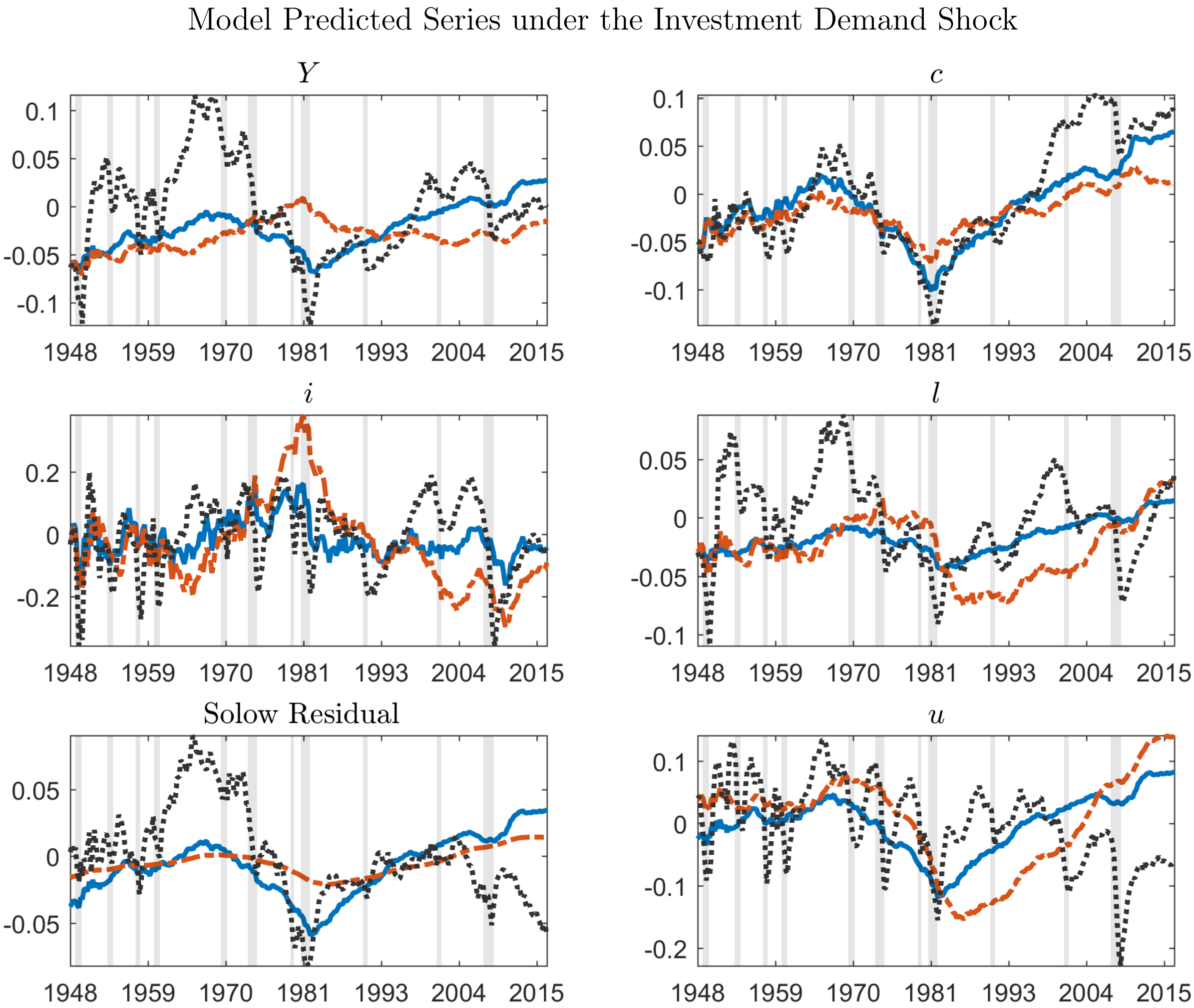

Figure 10: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Model Predicted Series under the Exogenous Expenditure Shock

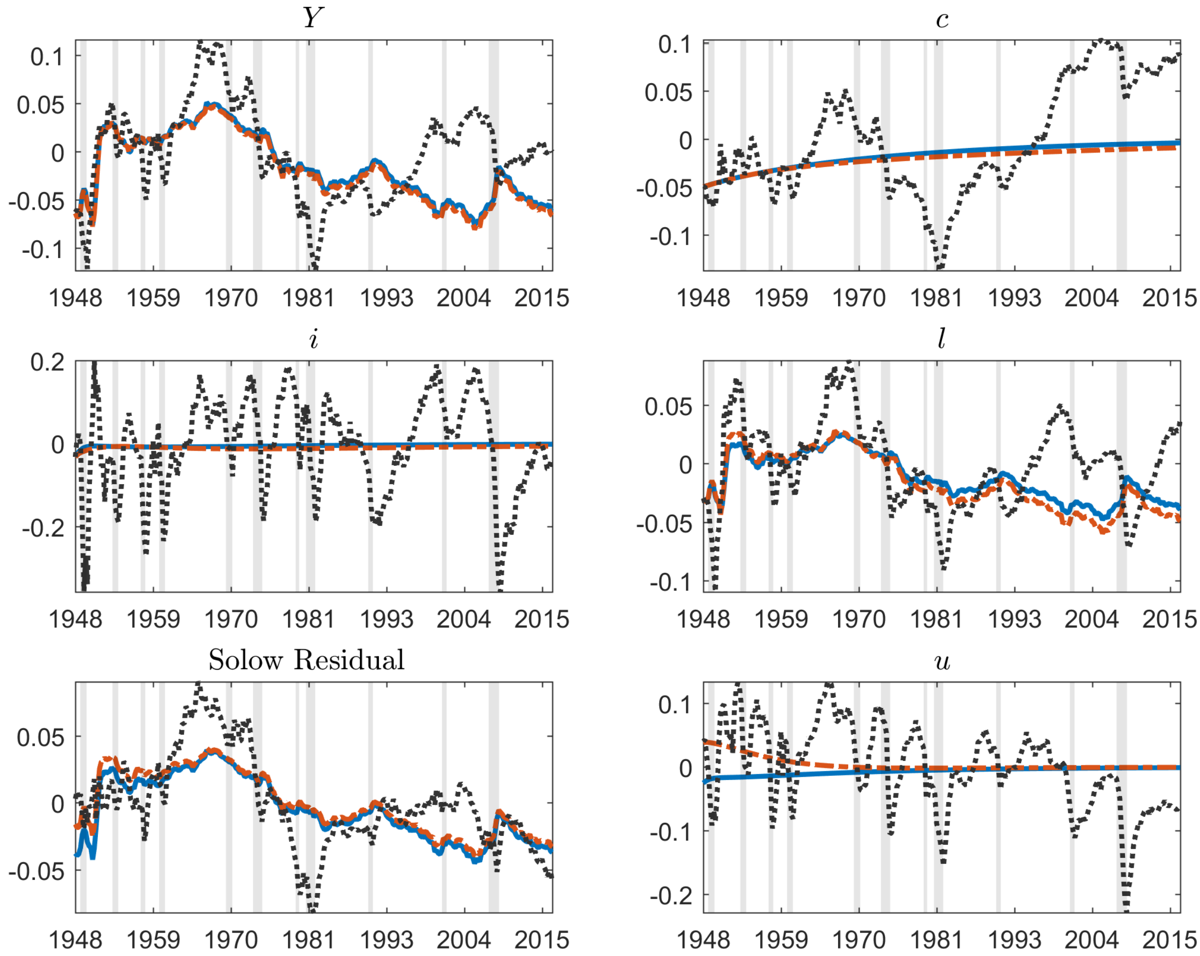

Figure 11: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Model Predicted Series under the Labor Productivity Shock

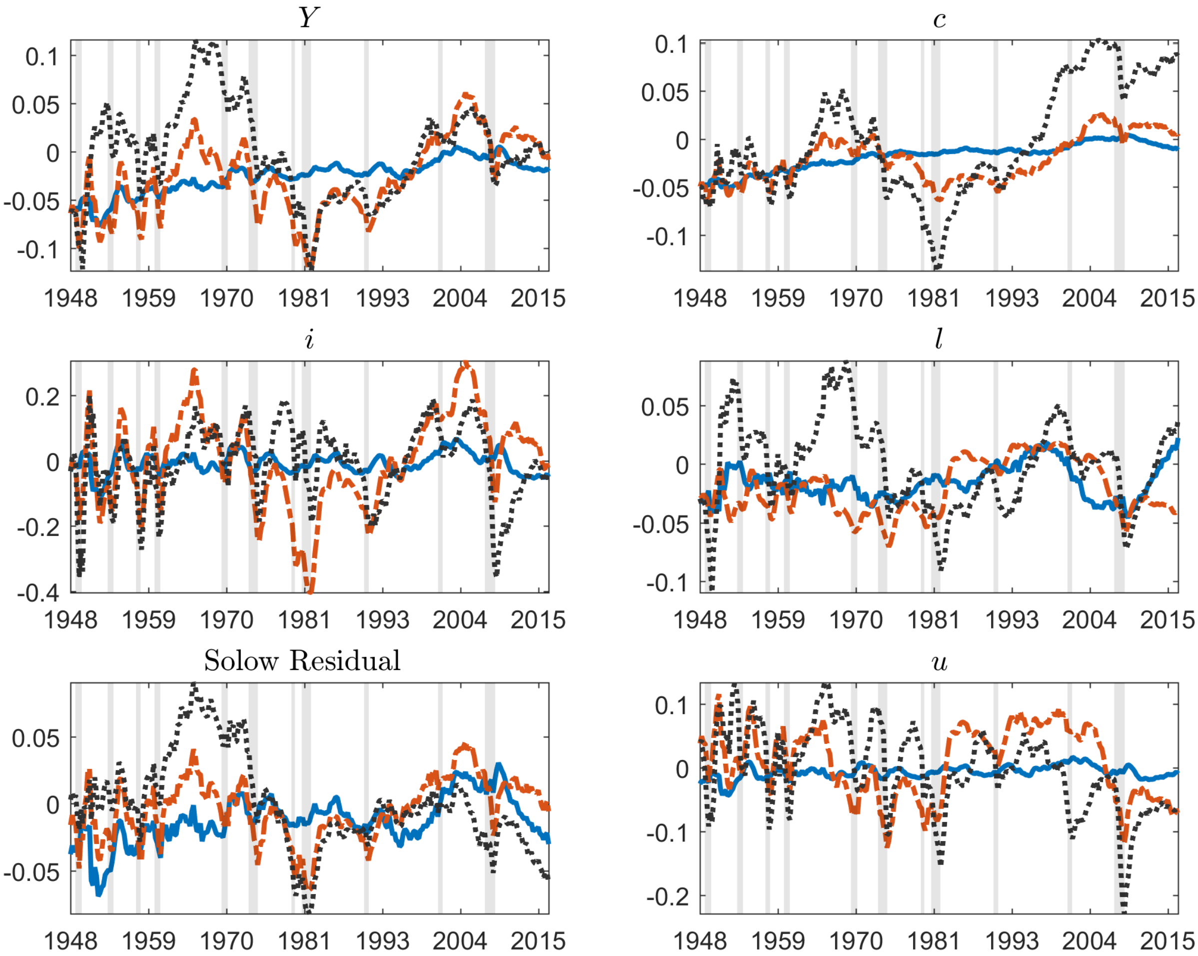

Figure 12: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Model Predicted Real Wage Rate and the U.S. Data

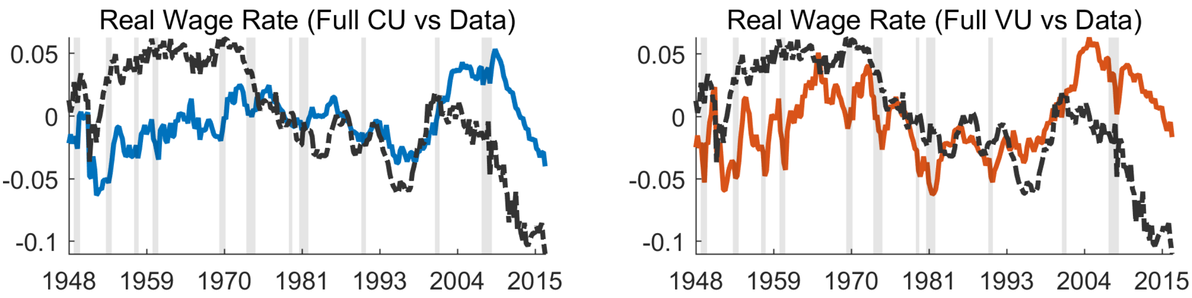

Figure 13: The solid lines are for the model predicted series. The dash-dotted lines are for the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA and the BLS. All variables are logarithms of the original series.